News in the Context of Regional and Functional Variation

–

A Corpus-Based Analysis of Newspaper Domains across Varieties of English

Von der Philosophischen Fakultät der Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des akademischen Grades einer Doktorin der Philosophie genehmigte Dissertation

vorgelegt von:

Jennifer Fest

Berichter:
Universitätsprofessorin Dr. Stella Neumann
Universitätsprofessor Dr. Thomas Niehr

Tag der mündlichen Prüfung:
23.11.2016

Acknowledgements

To begin with, I would like to thank my supervisor, Prof. Dr. Stella Neumann, for her constant encouragement and support over the last five years. I have always enjoyed the discussions we had, and hope that this work reflects how much I learned from them. Furthermore, I would like to thank Prof. Dr. Thomas Niehr for agreeing to be the second reviewer of this thesis and for his support for my research stay in Australia, and Prof. Dr. Sven Kommer for chairing the examination committee at my defence.

I am particularly grateful also to my former and current colleagues, Adjan, Aljoscha, Antje, Arndt, Caroline, Elma, Jessica, Julia D., Julia V., Nicole, Paula, Simone, Stella S., Sven, Tatiana and Timo, for their moral support and patience, as well as for their feedback on numerous presentation rehearsals, to the statistics wizards Bodo Winter and Stefan Evert for letting me participate in workshops whenever possible, and to my officemate Rebekah, for her endless patience and for listening to even more rehearsals without ever complaining. A big thank you also goes out to all my friends, for whom I did not have as much time as I would have liked to while writing this thesis, and who were always understanding.

This work would not have been possible without the input I received during my research stay at the University of Sydney. I am greatly indebted to Monika Bednarek for supporting me while there, and to all those who worked with me and took the time to give me their much appreciated feedback. My thanks also goes to the DAAD for funding me for this stay as well as for conference participations in Argentina and Spain.

Most of all, I would like to thank and dedicate this thesis to my family, my mum Brunhilde, my dad Helmut and my brother Dennis, for keeping me going with their unwavering support. I could not have done it without you.

Contents

Illustrations

Figures

Tables

List of Abbreviations

ACT	Australian Capital Territory
AUS	Australia
CDA	Critical Discourse Analysis
ECO	Economy
EFL	English as a foreign language
EIL	English as an international language
ENL	English as a native language
ESL	English as a second language
ESP	English for specific purposes
HK	Hong Kong
HN	Hard News
ICE	International Corpus of English
KEY	Kenya
L1	English as a first language
L2	English as a second language
LIFE	Lifestyle
NSW	New South Wales
NT	Northern Territory
OED	Oxford English Dictionary
POL	Politics
QLD	Queensland
SA	South Australia
SCMP	South China Morning Post
SFL	Systemic Functional Linguistics
SPO	Sports
TAS	Tasmania
UK	United Kingdom
USA / US	United States of America
VIC	Victoria
WA	Western Australia

1 Introduction

According to Ethnologue (2016), English is one of the most widespread languages of the world, and although both Spanish and Chinese count more first-language speakers, English is by far the dominant language when taking into account the numbers of people who speak it as a second or foreign language. Since its spread during colonial times, it has been established as an official language in countries all around the globe, and can be, and has been, called a "World Language" (e.g. Shaw 1981, 109, 118; Schmied 1989a, 7; Schneider 2007, 1). However, it is questionable whether a diversity of this dimension can still be observed from the perspective of one English language, or whether it is not more likely that individual varieties have, due to their own paths of development, drifted apart far enough to be recognised as separate standards or Englishes (cf. Görlach 1996, 153-54; Mesthrie and Bhatt 2008, 3).

This train of thought is reflected in the research in the field of variational linguistics, which looks closely at individual varieties of English and intends to describe them in great detail, looking at them in isolation as well as comparatively. Apart from defining particular characteristics of varieties, it is another focus of variational linguistics to define different types of varieties depending on the status the English language has in the respective society or culture and the way people are confronted with it as an official language as well as in their everyday lives. In this context, models have been suggested which aim at depicting these differences (e.g. McArthur 1987; Görlach 1990; Kachru 2006).

Although these models recognise and describe the different status of English varieties, the categorisations are rather broad and create many borderline cases which do not fit into these predefined areas. Furthermore, the characteristics of a language along which this categorisation is to take place are often very abstract and allow room for diverse approaches and groupings. The study at hand therefore assumes a different measurement for a variety's status which is based on more individual descriptions of varieties rather than a preset taxonomy.

According to Halliday and Hasan (1985, 38), language adapts to the uses it is put to. Analysing language in an isolated state neglects or even entirely ignores the influence that external, situational factors have on the way we use language: "The complete meaning of a word is always contextual, and no study of meaning apart from a complete context can be taken seriously" (Firth 1969, 7). This applies to the analysis of individual discourses and registers, but also offers a different perspective for defining varieties. Assuming that every situation in which language is used results in the development of a register characteristic for this very context, this, in reverse, means that the more registers a variety has formed, the more diverse are the contexts in which the respective language is used in the society in question. Furthermore, the degree of differentiation between the individual registers allows conclusions about the frequency of this particular use. If a situation occurs regularly or is highly normed in the society, its register will have formed strong linguistic characteristics that clearly reflect the context. In contrast to this, a register of a less prevalent context will most likely exhibit fewer specific features and thus a weaker distinction from other registers within the variety. It can thus be said that the number of registers and their degrees of specification are indicators for the developmental status a variety has (Mollin 2007; Neumann and Fest 2016). As it would be impossible to adequately cover a whole variety, let alone several in comparison, the study will be narrowed down to examining one of the most frequently encountered discourse types in our everyday lives, namely the language found in newspaper writing.

News in general is produced, and even more so consumed, by many different people, and the language used in this context is just as diverse as its addressers and addressees. Who is the intended target group of a news item as well as its main function heavily depend on the topic and the newspaper – political news and lifestyle items for instance aim at different readers, although the target groups may of course overlap. Similarly, their goals can vary from being informative to entertaining or even advisory. In general, news represents "snapshots of our life and our culture" (Reah 2003, 1) since they shape our perception of reality due to their portrayal of events, but at the same time underlie the principle of supply and demand to a degree that allows the readers' news preferences to influence the journalists' choice of topics. A newspaper that does not report stories readers are interested in will not sell, and therefore not have any economic future. The "snapshots", as Reah puts it, can be seen as including this circle and thus depict both the producers and the receivers of news, influencing each other by what they report and expect. While the readers' interest and hence the influence on the news production is relatively clearly represented in sales figures, however, the impact of news reports on the readers cannot easily be measured due to the unknown nature of the recipients. Researchers as well as journalists can only make assumptions about the target group, which then allows no definite conclusions on this part. Instead, every study working on news will have to focus primarily on what is known or can be included as independent or control variables, and analysing the language involved is the most effective way to understand how information is conveyed and which foci and values receive particular emphasis in a news story, as well as the medium's broader function in society (Beard 1998, 33; Bell 1998, 65).

1.1 Motivation and Aim of the Study

In this context, the study at hand aims at approaching the issue of different Englishes and their varying status from a functional perspective. While the role of English in native varieties is generally less disputed (cf. Kachru 1981, 15-17), it is particularly varieties from former colonial areas which are of interest here; in these regions, English was introduced very differently and language policies have varied greatly, which has resulted in different attitudes towards the language as well as discussions about its official status and use. Even though varieties might fall into the same category in a broad model of variation, differences with regard to their register formations and the use of English in varying domains can therefore be expected.

In order to cover a diversity of socio-economic as well as historical backgrounds, the present study includes varieties from two regions in which English is predominantly spoken as a second language, namely Kenya and Hong Kong. Whereas in the East African country English was introduced only marginally to selected workers and the colonial period ended violently (e.g. Abdulaziz 1991, 394-95; Mazrui and Mazrui 1996, 272; Skandera 2003, 11), Hong Kong experienced a more thorough spread of English and a more peaceful handover back to China, which, however, did not happen until 1997 (e.g. Pennington and Yue 1994, 2; Pennycook 1998, 1, 95-107). The divergent histories and self-conceptions have since led to very different directions in language policy; furthermore, the two regions of course differ greatly in demographic factors such as size and population, but also regarding the development of economy and infrastructure. Nevertheless, in most models of English variation the varieties from Kenya and Hong Kong are categorised similarly (e.g. McArthur 1987, 11; Kachru 1988, 5; Görlach 1990, 42; Schneider 2007, 133, 189), and will therefore be used as a test pair for a functional approach to variation.

To ensure that the results are adequately put into historical and social contexts, it is necessary to not only compare these varieties with each other, but also to native varieties that might have an influence on them. The origin of English in both Kenya and Hong Kong is British, yet nowadays, US-American English is present in many areas such as popular culture and politics, which makes it potentially influential as well. Especially for Hong Kong, Australia is another reference point as it is geographically closer and therefore holds an implicit news relevance due to proximity (Bednarek and Caple 2014, 155-56). Differences between the varieties of Hong Kong and Kenya do not necessarily have to originate from their own societies, but can be caused by an orientation along different native Englishes, which can be found only if these are included as well. The present study will therefore draw on data from the five regions of Kenya, Hong Kong, Australia, the UK and the USA.

This international coverage of English is even more relevant when taking a comparative look at news discourse. In order to reflect functional variation, which is the basis for determining the status of the varieties, the dataset used here contains newspaper articles from five topical domains, economy, politics, sports, lifestyle and hard news. These areas are predefined in most newspapers and can therefore be expected to show differences in all regions; to avoid mis- or over-interpretation of the results, it is essential to compare these differences across the regions and find domain-dependent universals before drawing conclusions about the varieties.

On the basis of these theoretical assumptions, the study aims at answering two major research questions which reflect the perspectives included in the dataset. Firstly, the newspaper domains will be analysed comparatively to trace potential universal characteristics and differences between them and see which domains are related and which ones are rather distinguished. The results of this comparative direction will be put into relation with the five varieties and lead to conclusions about the newspaper discourse in the respective societies. Secondly, and drawing on this, the developmental status of the varieties, particularly Kenya and Hong Kong, will be defined with regard to the functional variation they display in this discourse. The insights gained here will be placed in the contexts of and compared to the previous categorisations of these varieties to see whether the functional picture they display can enrich these prior definitions, or whether they contradict them.

With the combination of these two perspectives, the study contributes to both media analysis and variational linguistics, but also addresses methodological issues that arise from such an approach. The register model described within systemic functional linguistics (Halliday and Hasan 1976, 1985; Halliday and Matthiessen 2014; cf. also Lukin et al. 2011; Wegener 2011, 104) will serve as a basis for the analysis as it defines three parameters for a contextual situation, and thereby as influential on the language in use, namely the field, tenor and mode of discourse, reflecting the topic, the participants' relation and the language role. This tripartite approach offers the possibility to find differences in language use not just in general, but to trace them back to the aspect of the situational context from which they originate. In order to apply this framework adequately, however, and make use of this potential, concrete linguistic features that reflect these dimensions have to be defined. Such operationalisations have been conducted in previous studies (e.g. Steiner 2004; Neumann 2013), yet are hardly ever concurrent. The present study will elaborate on these steps to define the linguistic parameters for the analysis of newspaper language and also discuss the functions of linguistic features in this particular discourse.

1.2 Thesis Organisation

Since the investigation combines three different research fields – media studies, variational and systemic functional linguistics – the next chapter will describe the theoretical foundations from these areas and put them into relation with each other. The first aspect to be discussed is the expansive field of English as a global language, which has developed very diverse approaches and ideologies depending on political as well as scientific factors. The dominant strands and major milestones will first be described in general, before the various sets of terminology which have emerged in the context of this field will be entangled and defined for the present study. In section 2.1.3 the focus will then turn to a closer description of the L2 varieties relevant for the analysis and give an account of the development of English in Kenya and Hong Kong. As a last step, the prevalent research foci and methods applied in this field will be discussed to define prior insights and gaps.

A similar structure will underlie the description of the functional perspective. Newspaper articles are, in accordance with all major definitions, to be seen as texts, yet different traditions and approaches to text analysis can be found which represent different emphases and research goals. The second part of the theoretical background will therefore again start with a broader perspective, comparing the dominant approaches and their developments as well as defining relevant terms and concepts. The functional approach, which forms the basis of the present study, will then be taken up in a second step and described more closely, before major research themes in the field will be identified and put into context with the current goal.

The third part of this chapter will then deal with the broad area of media studies. This field is, in its development and origin, less linguistically influenced than the other two, and will first be put into its overall context before the focus is shifted to linguistic perspectives that have been taken to media language. Here, only general trends can be described however as a vast number of studies working on media language take an empirical approach and are based on samples from concrete media producers and countries, providing precise, but specific results. To keep the focus on the present regions, section 2.3.3 will therefore summarise the insights gained so far for the varieties relevant here, giving a state of the art of research conducted in each of the five cases. Furthermore, the different newspaper landscapes will be described in detail in order to give an overview of the situations and define the context in which the data used in this investigation was produced and has to be interpreted.

Having defined the theoretical background and context for the study, the succeeding chapters will be concerned with methodological aspects of the analysis. As was mentioned earlier, the operationalisation of the linguistic description within systemic functional linguistics is not entirely straightforward, but requires reflections on the discourse under investigation here as well as on previously defined strategies. Chapter 3 will first give a brief introduction to quantitative approaches to variation before describing theoretical and methodological implications for each contextual parameter – field, mode and tenor – and defining the concrete linguistic markers to be analysed for this study. Chapter 4, then, will turn to the dataset on which the study is based and define the target population and sampling frame as well as discuss the issue of representativeness in corpus design. Section 4.2 will describe the processing of the data, including linguistic and non-linguistic annotation.

Within this defined scope, chapter 5 will then present the analysis and describe the results in detail. The structure will be mapped onto the three aspects of field, mode and tenor of discourse and their respective subdimensions, and domains and varieties will be described together for each parameter so that differences and similarities in the two directions of comparison are immediately visible. Detailed results will be discussed when presented, before chapter 6 then summarises the significant trends and discusses them in relation to variational and media studies, bringing them back from fine-grained, individual findings to the more extensive level of the research questions. Chapter 7 will conclude this study by revisiting the main assumptions and questions as well as the method applied here, and give an outlook on potential future research aspects and questions.

2 Theoretical Background

The aims of this study as presented in the introduction require the combination of three different scientific fields.

First, the function of English on a global scale has to be considered in order to properly understand the development of different varieties. In this context, the process of its expansion around the world has been discussed and analysed from various perspectives, and has itself been influenced by societal and political changes both in the language's 'home country', England, as well as its new territories. The thus created history of English has countless dimensions depending on contact languages, cultures and ideologies, all of which are reflected in the research and the trains of thought that led to the understanding of English as a global language and the acceptance of varieties as legitimate forms of it. These developments will be traced in section 2.1; furthermore, out of the five varieties included in this study the two cases of Kenya and Hong Kong will be described in more detail and serve as examples of language development under very different socio-political circumstances.

Secondly, an overview of possible approaches to text analysis and traditions in the field will be given in section 2.2. Newspaper texts are very particular with regards to both structure and language, and it is important to reflect on different definitions of text and the various perspectives on newspaper articles they open up. As the discussion will show, there is no 'right' way of analysing news items, but every approach lends itself to different research questions. The systemic functional framework, which was chosen to form the basis of this study, will be then be explained in more detail in sections 2.2.2 and 2.2.3.

The third perspective to be considered when laying the foundation for this work is that of media studies. Although this analysis has a linguistic focus, it would be negligent not to take the origin of the texts into account and draw on research that has been done in this field. Section 2.3 will therefore give a brief description of the development of media studies, including sociological work, but also journalistic points of view. Since the five varieties discussed here vary greatly in terms of mass media coverage, we will then take a closer look at all five cases to provide the background knowledge against which the results of the analysis are to be interpreted.

All the fields treated in the following sections have accumulated a wide range of terminology, not all of which is always used consistently. Terms relevant for this work will therefore be defined as clearly as possible in the respective sections. Furthermore, the most common methods of analysis in each field will be identified in order to grasp the foci that have so far been dominant in the research and put the current work into context.

2.1 English in a Global Context

As early as 1981, Gerry Abbott wrote in his editorial to the first ever issue of the journal *World Englishes*:

> English is now above all an international language, used or understood in most parts of the globe and in a great variety of circumstances. In today's world it serves as a means of expression and communication not only among people who have acquired it as their mother tongue but – on an even more remarkable scale – with and among non-native users, whose mother tongues are many and various. (Abbott 1981, 1)

In this brief but exact introductory statement, Abbott touches upon several issues that are crucial for researching the English language and are, despite continuous efforts and insights, becoming ever more complex as well as relevant in our modern and fast-changing world.

Not just since the publication of Abbott's editorial, but for most of the last centuries and especially since the era of colonialism, the number of people who speak English has increased steadily. In 1962 Quirk, estimating due to lack of data from many parts of the world, gave the numbers of 250 million people with English as a mother language and another 100 million speaking English as either second or foreign language (Quirk [1962] 1968, 8). In 1985, he had already raised these figures considerably: in a conference presentation on "Progress in English Studies" he calculated about 300 million mother tongue users, 300 million second language speakers and 100 million for whom English was a foreign language. Less than a year later, these numbers were again challenged by Crystal. Drawing on the work of Gunnemark and Kenrick (1983), he lists 60 countries and "various British territories" (Crystal 1985, 8), estimating the speakers of English as a first or second language at a total of about 300 million and 1.3 billion respectively, only to raise these numbers again two years later in his *Cambridge Encyclopedia of Language* (Crystal 1987). Finally, he arrives at more than 1.7 billion potential[1] English speakers, not even counting those who

[1] In the *Cambridge Encyclopedia of Language* (1987), Crystal assumes a total of 350 million mother tongue-speakers and estimates the number of users of English as a second language to be around 1.4 billion, which corresponds to the population sizes of the countries where English functions as an official

learned it as a foreign language. In order to sort these cases properly, the empirical data would have to be very exact and reliable, an aspect that Putseys (1993, 4) declares to be a major problem: "These numerical data [concerning language usage] must be viewed with the same caution as any other type of demographic statistics: their accuracy depends on the effectiveness of each national system for registering them."

Probably because of this, many other studies have abandoned the mere counting of speakers and have instead turned to view English as a global phenomenon and analyse it more closely in this context. This chapter will follow this trend and, in the succeeding sections, will give brief recounts of the theoretical concepts of *World English* and *World Englishes* as well as the categorisation of speakers of English as a first, second and foreign language. It will then narrow the scope to what has been termed *New Englishes*, the field of study to which Kenyan and Hong Kong English belong. In a last step, a brief summary will be given of methodological approaches and theoretical perspectives which have so far been applied and studied with regard to English as a global language.

2.1.1 World English and World Englishes

One does not have to know the exact number of English speakers on the globe in order to work with the assumption that there are many millions and that English is therefore a world language. Suggestions as to what that might mean and imply differ greatly however, and the ensuing discussions have revealed just how far-reaching and obscure the concept of English really is.

As Bailey (1985, 3; cf. also 1992, 93-94) rightly points out, "the idea of a world language is one of considerable antiquity". He gives the example of the story of the Tower of Babel from the Book of Genesis as the first instance of mankind wishing for a universal language, and ending in distress because of losing precisely that. Old as that story might be, the idea of one shared language spoken by everybody can be found recurrently in linguistic literature. The 17th century saw numerous attempts at creating philosophical languages, constructed languages that are based on the sim-

language. A certain margin of error should thus be allowed to account for people who, despite the official status of English, do not speak it as a second language.

plest rules and are therefore easy to learn and spread (see e.g. Eco 1995, ch. 15; Seargeant 2012, 81-87). Two famous works, among many others, were presented by Francis Lodwick in his books *A Common Writing: Whereby two, although not understanding one the others language, yet by the helpe thereof, may communicate their minds one to another* (1647) and *The Groundwork or Foundation laid (or So Intended) for the Framing of a New Perfect Language and a Universal Common Writing* (1652) (cf. Eco 1995, ch. 13). In the late 19th and early 20th century, the issue was again raised rather radically; in 1888, Bell declared that a global language was inevitable and that English was destined to fill that gap, although in a less complex form than was commonly spoken: "Let a committee be appointed", he wrote, "consisting of one British and one American member, to investigate the subject, and suggest such changes as would remove anomalies, and I feel convinced that they would readily create a new and simple tongue in the form of what may be called 'world English'" (Bell 1888, 40). In 1910, Braendle presented his own world language, "Veltlang", which was to be based on English in grammar and lexis, but concerning phonology and orthography was to be fashioned after "the most primeval characters […] ever used by the human race" (Braendle 1910, title page). After him, Hamilton (1928), too, suggested a simplified version of English as a worldwide language. The most radical and, from a linguistic perspective, frightening notion however, was published by Alderson in 1908; his suggestion is nothing short of creating a monolingual world: "The correct policy to follow is to kill all the languages […] and substitute for them one only: English" (Alderson 1908, 160). He claims that not only would this save all the money and time "wasted" (ibid., 72) on learning other languages, but it would put an end to wars and racism and finally create world peace. Apart from being imperialistic, this chain of reasoning is of course optimistic to the point of naivety – and, fortunately, is not taken up again in the literature dealing with English as a world language.

The idea of a world language lingered on however, and another, although a much less radical suggestion of that kind was made by Quirk (1981), who suggested "Nuclear English" as a world language. Quirk states that it would have to be "decidedly easier and faster to learn than any variety of natural, 'full' English" and that "since nothing should have to be 'unlearned', the lexical and grammatical properties of Nuclear English must be a subset of the properties of natural English" (ibid., 155-56).

Independent of their level of radicalness, these approaches were positioned around the central notion that one language should be imposed on everyone, and that this would solve all communication problems in the future. They underestimated or failed to consider entirely several major consequences however; on the one hand, speakers of English on a native or advanced level would face the problem that they would have to downgrade their language whenever communicating with someone who learned only a simplified version of English, a behaviour which would be impossible to enforce. On the other hand, the factor of language change and the tendency for languages to develop differently depending on where and by whom they are spoken would very likely make every effort to establish and retain a world-wide language fruitless. There is of course no empirical data to this effect,[2] but evidence from the histories of other languages makes it more than likely that an artificially created world language would remain homogenous only for a very short time. It would start to change the moment it was introduced, being adapted to social groups, formed to suit societies' needs and influenced by surrounding languages, ideologies, and geography (e.g. Strang 1970, 17; Kachru 1985, 22; Eco 1995, 292; Widdowson 1997, 142; Seargeant 2012, 26). In other words, it would form regional and functional varieties – something for which no artificial language is needed, as English is doing just that on a global scale already.

Recognising this, the research field of World English experienced a slight shift from the attempts to establish (a version of) English on a global scale to analysing the English that was already there. This of course presupposed accepting the fact that English differed greatly depending on where and by whom it was spoken. In his "History of the English-speaking people" (1956-1958), Churchill acknowledges this diversity within the English language, but regards the varieties as nothing more than slight alterations of the English spoken in England. Although the Oxford English Dictionary (OED 2016) records the first instance of the word "Englishes" in the sense that it refers to regional varieties of the language English in the year 1910,[3] it was not before

[2] The only constructed language for which data is readily accessible is Esperanto, a language developed by Zamenhof between 1872-1885. Intended as an auxiliary language to simplify international communication, its number of speakers is estimated by Ethnologue at about 2,000,000 (Ethnologue 2016).
[3] This first occurrence originates from a newspaper column written by the American journalist Henry Louis Mencken and published in the *Baltimore Evening Sun*. The column bore the headline "The Two Englishes", and discussed the differences in spoken language between British and American English. The OED names other occurrences of *Englishes* before this, which have different meanings. As early as in the Old English period, the word was used to refer simply to English people. In the 17th and 18th century, it was recorded to mean either translations from or to English, or individual English words in foreign texts. Also after Mencken's column the term was used in different ways, mainly referring to differences between spoken and written English (OED 2016).

the mid-20th century that the term became more frequent in scientific texts (cf. Mesthrie and Bhatt 2008, 3). A simple morphological analysis is enough to unveil the basic idea behind this new use – as Chevillet (1993, 32) explains:

> When we transform 'English' (N1) into 'Englishes' (N2) we change a 'mass (or 'uncountable')' noun into a 'count (or 'countable')' one, resorting to the well-known process of *functional shift* (or *conversion*, or *zero derivation*). It is the same when we speak of 'Italian *cheeses*', 'English *cheeses*' or 'French *cheeses*': we are referring to different kinds/varieties of the same (though diversified) 'thing', i.e. *cheese*. (emphasis in original)

The birth of this concept, logical as it was, triggered mainly one question however, namely that of the future of the English language. Both Bailey (1985) and Maley (1985) painted two possible scenarios for this: either, English would break up into various, eventually no longer mutually intelligible languages, like Latin did (see also Schneider 2007, 68); or it would remain united due to the fast exchange of communication and information made possible by modern technology. Eventually, Maley settled for a compromise: “What I foresee then is the continuing development of local or specialist dialects, but within a framework of an internationally understood variety of English” (1985, 32). This prediction raised other issues which he, too, was quick to add: “Where then will ‘real’ English be spoken? What will be recognised as the ‘best’ kind of English?”

These questions clearly show that, despite the insight that there were different varieties of English, the overall perspective was still fairly centred on one core language, located either in the USA (as suggested e.g. by McArthur 1987) or Britain (see Christophersen 1988). Strevens’ (1980, 86; 1992, 32) model “Global family of Englishes” (Figure 2.1) creates a visual representation of this by splitting English into these two varieties and then drawing branches from them across a map of the world, indicating which Englishes come from which English.

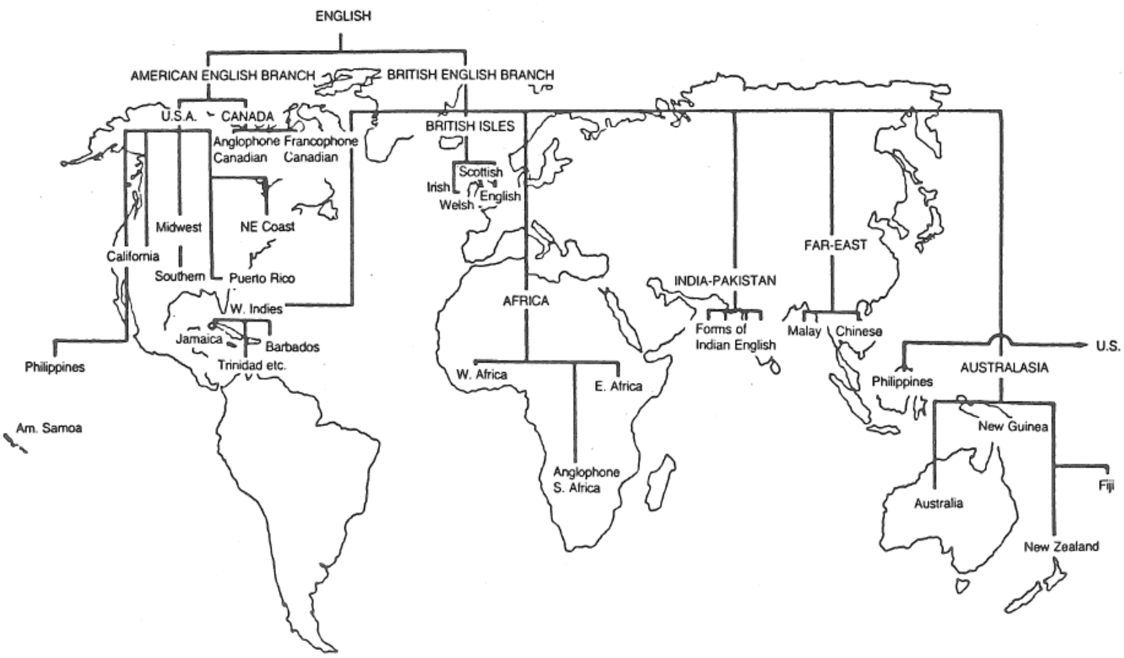


Figure 2.1: "Global family of 'Englishes'"

Strevens was the first to create a model which took synchronic variation into account in more detail. Up to that point, models had focused on a diachronic perspective, either representing the historical periods of English, or working with the image of family trees, thus depicting the origins of English in the Germanic language family, usually going back to the assumed language of Proto-Indo-European (cf. McArthur 1998, 80-93). Neither of these two types allows much room for synchronic diversity, a point which McArthur (1992, 13-16) and Chevillet (1993, 30-32) rightly criticise. Strevens' model attempts to combine the current state of language relations with the historical origins of the varieties, but it suffers from more or less major flaws in both aspects: concerning the diachronic perspective, the image of the family tree implies the notion that after 'giving birth', the parent will, at some point, die. This, to put it bluntly, is how nature works, but it is not a suitable metaphor for the dynamics and lives of languages (Chevillet 1993, 32). This problem of visualisation, however, could be easily overlooked, were the model capable of representing the rather flexible state of English in its current state. The image of the world map has advantages over many other designs, as it implies without saying that there are regional and social dimensions to be considered when researching English. Yet it fails to express the existing and potential interconnections between varieties other than British and American English as well as between varieties and other contact languages. It allocates no space to these linguistic microcosms and thereby appears too static and nonexpendable for a language which has affectionately been called "the living laboratory of World English" (McArthur 1987, 10).

What can be learned from the example of Strevens' model is that a combination of diachronic and synchronic perspectives in one design almost inevitably leads to shortcomings in both. Each of these aspects is very complex in itself, and giving justice to just one of them in a simple diagram is already a feat. Furthermore, the prominent status given to the varieties of British and American English, although reasonable from a diachronic point of view, appear unproportional looking at the current state of English. This insight did not only pose a problem for visualisation however, but was first of all one of acceptance. McArthur (1987, 9; 1998, ch. 3) points out what he calls "the anxiety among many that the great historical monolith of the English language has begun to crack", an observation which is shared by Christophersen (1988, 16). In his own, mainly synchronic model, McArthur (1987, 11) therefore refrains from placing any regional standard at the centre, and instead creates a wheel structure with the abstract idea of a "World Standard English" as the hub (see Figure 2.2).

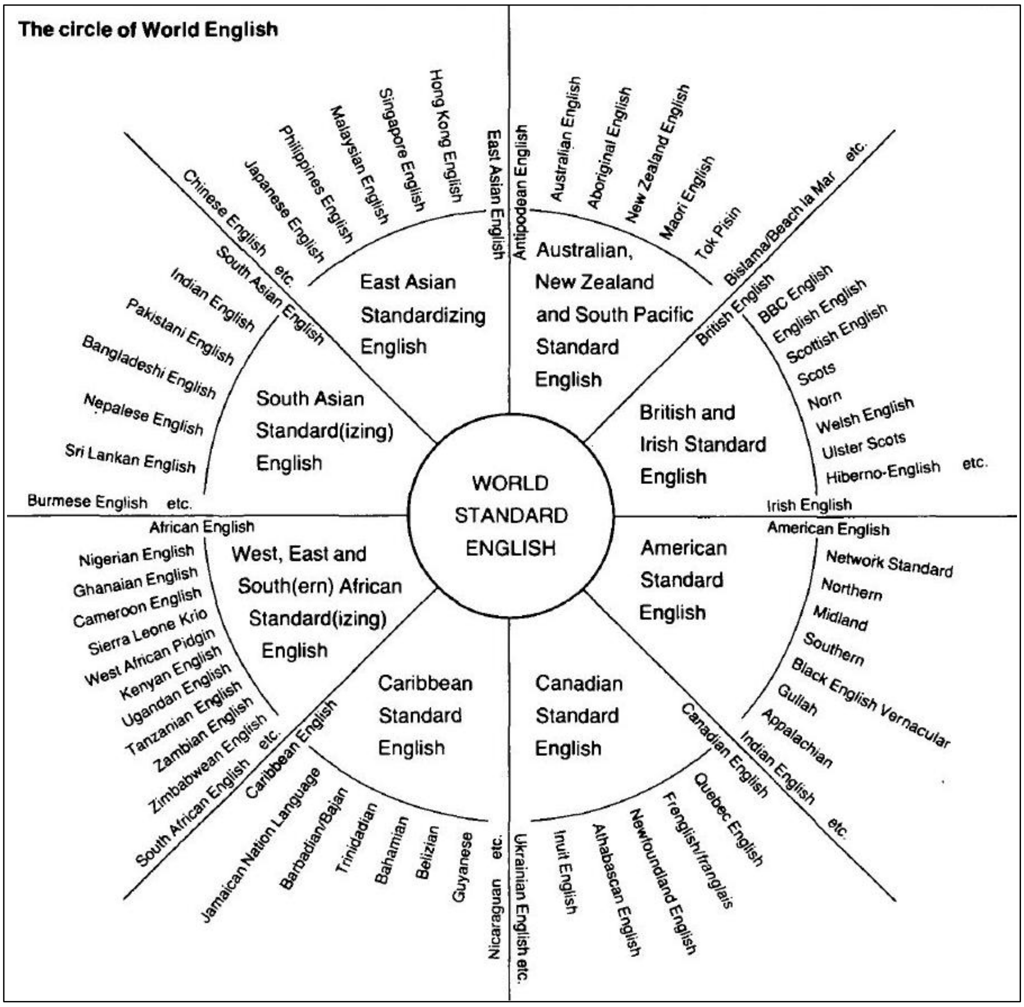


Figure 2.2: "The circle of World English"

This representation is much more open to possible expansions in the language development, as spokes could be added to create new sections in the wheel and there is no outer limit. In this sense, the model can be said to have a diachronic perspective in the making, as future changes can be embedded without changing the general structure. A similar approach was taken by Görlach (1990, 42) in his model published three years later (see Figure 2.3). Again, one standard, in this case termed "International English", forms the centre of the wheel, being followed by regional or national standards in the next ring, sub-regional "semi-standards" in the third and dialects and "non-standards" in the last. Outside his wheel, Görlach places pidgins, creoles and related languages, thus, like McArthur, reserving room for developments to come.

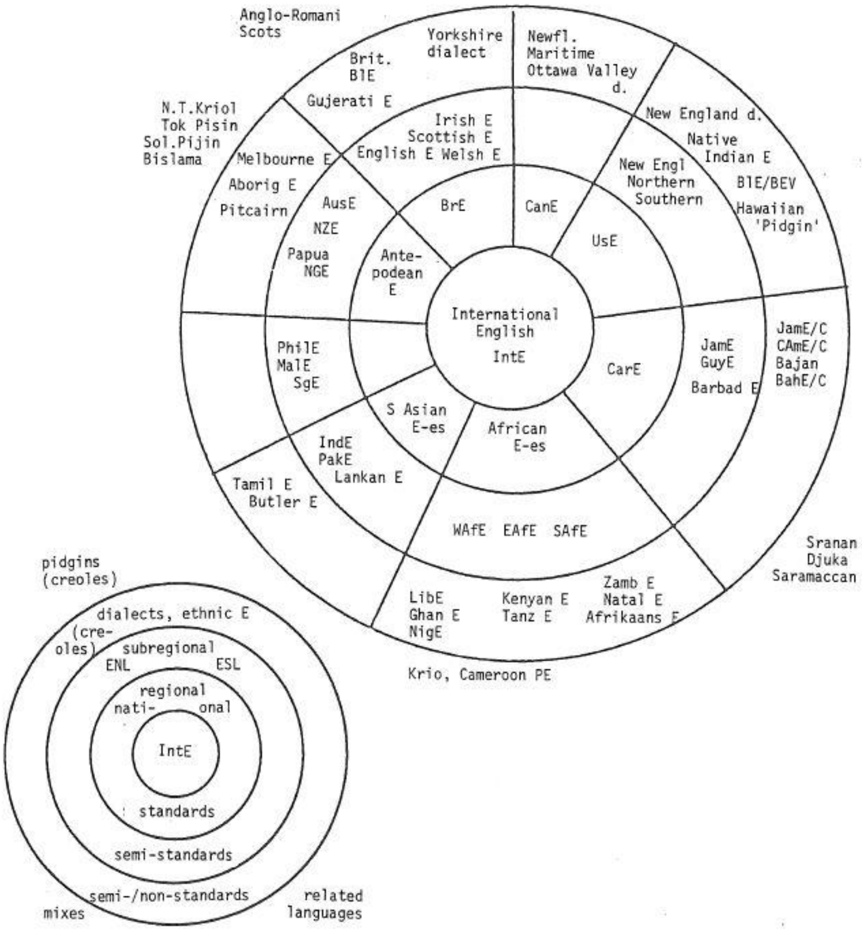


Figure 2.3: "The status of varieties of English and related languages world-wide"

Despite all far-sightedness, the models do not seem to work without the idea of some abstract 'macro-English' as a point of reference (cf. also Mesthrie and Bhatt 2008, 27-29). The first attempt at modelling the English language without such a fixed anchor was made by Kachru in his "Three concentric circles of English" (Figure 2.4).

While the first (1985) paper in which the model was mentioned featured only a description of the concept, later publications included a visualisation. The model essentially displays three circles: the smallest “inner circle”, the medium-sized “outer circle” and the largest “extending circle” (which was renamed “expanding circle” in a further publication of the model in 1992). The varieties in the inner circle are said to be “norm-providing”, the ones in the outer circle “norm-developing”, and those in the expanding circle are “norm-dependent”.

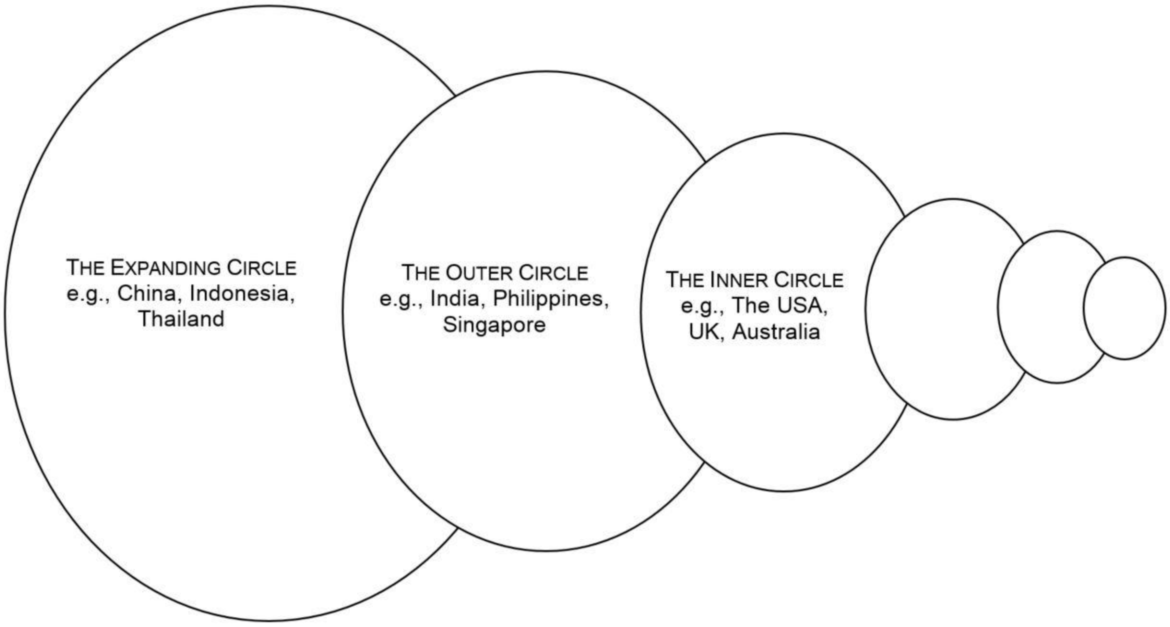


Figure 2.4: "Three concentric circles of World Englishes"[4]

This model was published in various of Kachru’s works (e.g. 1988, 5; 1992, 356; 1996, 71; 2006, 196), and has often been slightly altered. The basic idea behind it has remained constant however, and, in contrast to the preceding models, does not draw on an international variety of English. Like the wheels of McArthur and Görlach, the concentric circles contain space for future developments in the English language – in some publications, like the one from which Figure 2.4 was adapted, it even contains entirely empty circles to demonstrate that it is not only likely for varieties to enter existing circles or shift from one to the other, but that entirely new dimensions of the English language might emerge as well. Kachru’s model is now more than 30 years old, and already one might argue that particular spheres like for instance social net-

[4] This version of Kachru’s model was adapted from Kachru (2006, 196). In the original, the orientation of the model is vertical and the circles are placed beneath each other, with the largest one on top.

works, forums or online multiplayer games are developing their own varieties of English. These are neither dependent on any physically existing region nor is there any reliable way of determining the number of users or speakers. Kachru's solution to leave blanks and open up this space for yet unknown factors is therefore a very reasonable one. The distinction in the model based on countries on the other hand causes a problem: while varieties that are independent of regions might fit into the empty circles, differentiation within a region, especially based on social varieties such as African American Vernacular English or Asian American English, is not mirrored in the model (cf. Mesthrie and Bhatt 2008, 30).

Apart from this conflict, the model implies another major statement, namely that the outer circle-varieties do not just rely on the inner circle for their language, but develop norms of their own. This aspect is strongly emphasised by Kachru, and throughout his publications (e.g. 1985, 1991) he argues for a distinct autonomy of the outer circle-varieties and suggests that in the countries belonging to the outer circle, the local varieties should be institutionalised as the languages used in education (cf. also Modiano 1999, 22; Bolton 2006, ch. 2). His perspective was opposed by Quirk (1990; Quirk et al. 1985), who, although registering and accepting the notion of different varieties, related back to the fear mentioned above that the lack of a supra-standard would eventually lead to mutually unintelligible Englishes. To prevent this, he recommends for "non-native teachers to be in constant touch with the native language" (1990, 7) so as to spread major standards of English and override local tendencies. From a cultural point of view, this is of course fairly drastic: as Widdowson (1997, 140) in another paper rightly says, "the spread of a […] language implies adaptation and nonconformity". There are too many local influences in the different countries in which English is spoken for it to remain static. Rejecting his earlier concept of nuclear English, then, Quirk's idea of a common core seems much more realistic:

> A common core or nucleus is present in all the varieties so that, however esoteric a variety may be, it has running through it a set of grammatical and other characteristics that are present in all the others. It is this fact that justifies the application of the name 'English' to all the varieties. (Quirk et al. 1985, 16)

The concept reoccurs in the literature under different terms; Chambers (2004) and Sharma (2009) talk about "vernacular universals", and Kortmann and Szmrecsanyi (2004) specify "angloversals". In this sense, a shared 'Englishness' is not only logical, but also observable in empirical studies, as for instance Nelson (2006) in a large-scale analysis of the International Corpus of English (ICE) and numerous other studies following his example (e.g. Xiao 2009; Yao and Collins 2012), have demonstrated.

The scenario predicted by Maley and Bailey in 1985, namely a mutual intelligibility of all Englishes due to modern communication, therefore seemed to become a reality.[5] To distinguish it from the idea of World English, this concept is mainly referred to as "English as an International Language" (EIL), which is essentially a functional variation of English and serves the purpose of communication in such matters as economy, politics, science, and education, but also across regions (Matsuda and Friedrich 2011, 333; Seargeant 2012, 87-89). It does not equal the concept of English as a lingua franca, since the latter is usually meant to imply that at least one participant of the communication does not speak English as a first or native language, and English thus functions as "a means of communication between people who do not share a common language" (Seargeant 2012, 191). EIL includes this possibility, but goes beyond it.

To what degree an individual might need EIL can of course vary; Smith (1992, 75) argues that "it is unnecessary for every user of English to be intelligible to every other user of English. Our speech/writing in English needs to be intelligible only to those with whom we wish to communicate in English". Looking at today's communicative networks and demands however, the need for as well as existence of this facet of English can hardly be denied. Its basis is the common core described by Quirk et al. (1985); to really function as a world-wide language in a number of settings however, it has to go beyond that and thereby logically include linguistic features that are characteristic of some varieties, but not of others (i.e. leaving the common core). Depending on the setting, these additional features vary, eventually leaving us with not one EIL, but several world-wide, yet functionally specific, registers. This trend is reflected in the scientific literature as studies analyse functional varieties such as academic writing (e.g. Hyland 2002; Alméciga and Evans 2014; Hützen 2016; Peacock 2015), social media English (e.g. Raguseo 2010; Zappavigna 2011; Gillen and Merchant 2013), business communication in international settings (e.g. Böttger 2007; Du-Babcock 2013; Barés and Llurda 2013) or legal English (e.g. Thomas 1985; Bhatia 1993).

[5] When discussing intelligibility, it should be kept in mind that it is not a given feature, but requires the willingness of all discourse participants to understand and to be understood (Verschueren 1989, 50). It therefore also touches upon the aspect of speaker attitudes and attitudes towards accents, dialects and varieties (see Mufwene 2001, 122-23).

These registers vary with regard to their divergence from the common core – which enhancements are made to the basics depends on the function as well as the influence on the field of individual varieties. This differentiation of registers of course restricts the mutual intelligibility of all Englishes to a certain degree; particularly with regard to their lexicon, functional varieties are often specialised to a degree which renders them unintelligible to outsiders (Widdowson 1994, 383; 1997, 144).

It can thus be said that while the term 'World Englishes' refers to regional varieties of English, 'Englishes as International Languages', as it would have to be termed, describe functional variations of the language or the so-called 'English for Specific Purposes' (ESP) (see Widdowson 1997, 144). This distinction, although helpful from a terminological point of view, does not imply that the two categories do not influence each other – quite on the contrary, they steadily interlace and shape the landscape of Englishes. Taking into account both functional and regional variation settles, at least temporarily, a long-lasting debate about the nature of the international status of English, but only to let the floor be taken by another issue which was already touched upon above, namely that of norms and rights of English. After all, despite all research efforts and discussions, "the myth of a single 'golden' or 'pure' form of English dies hard" (Strevens 1981, 3). The next section will therefore look more closely at the ideas of nativeness and English as a foreign or second language, and define the meanings of these concepts for the purpose of this work.

2.1.2 EFL, ESL and the Concept of Nativeness – "The Empire strikes back"

When describing models like the ones presented in the previous chapter, it is inevitable to work with terms such as *native speaker*, *mother tongue*, *foreign language* or *second language*. Very often, the meanings of these terms are taken for granted, and are not clearly defined. Statistics about language users, too, usually work in these categories, as could be seen in section 2.1. In the literature concerned with the terminology, however, the meanings of these terms are far from clear. Furthermore, with people becoming ever more mobile and linguistic biographies ever more diverse, new terms arise to describe new phenomena concerning language use.

To begin with, the frequently used idea of *nativeness* should be treated with a certain caution. Not only is the notion itself a rather vague one, but it is also often used synonymously with other concepts, especially those of *first language* or *L1* and *mother tongue*:

> *Native* must surely be one of the most misleading and confusing terms ever employed in technical or semi-technical argumentation. Two other terms, partly synonymous with it, run it a close second or third: *mother tongue* and *first language*. A person's mother tongue is not necessarily his mother's tongue; nor is his first language always that which he learnt first, because *first* can mean 'first in importance' as well as 'first in time' and, alas, the two meanings are sometimes run together as if they were one and the same. All three terms, moreover, are surrounded by an aura of mystique and are heavily loaded with emotional connotations inimical to sober scientific investigation. (Christophersen 1988, 15; emphasis in original)

Native language and *mother tongue* are used synonymously almost all through literature to refer to "the language that one acquires from birth" (Seargeant 2012, 196; see also Mesthrie and Bhatt 2008, 4), and will be treated as such also in this study. The concept of *first language* however is defined rather unsatisfactory in most cases. Seargeant (2012, 192) explains it as "the initial language an individual acquires, or the language in which he or she is most proficient", thus effectively offering two definitions, one of which equals that of *native language*. Schneider (2007, 17) is of more help in this question, as he draws a clear line between the terms: L1-speakers of English to him are "speakers who, after having acquired an indigenous mother tongue, have sooner or later shifted to using English only or predominantly in all or many domains of everyday life".[6]

Despite this much-needed distinction between *first* and *native language*, it causes a problem in the next step, namely the differentiation of *English as a second language* (ESL[7]). This category is usually defined similarly to Schneider's explanation of L1 (see e.g. Crystal 1985), and yet a distinction can be made in the dominance of English. Widdowson (1997, 144) emphasises the official character of a second language (L2) by stating that "in so-called second language countries, English is used more widely for institutional purposes". This is taken up by Seargeant (2012, 192), who elaborates: "[ESL refers] to the use of English in countries where it has some official status (mostly due to the legacy of colonialism), or in which it is the predominant means of communication and is being learnt by people […] from non-native speaking backgrounds". English, then, is actively learned and in most cases an obligatory school subject due to the function of the language within the country in question. In contrast to English as a first language however, ESL is not necessarily dominant in a speaker's

---

[6] To distinguish it more clearly from English as a second language, Gupta (1994) coined the term *English as a dominant language*. As this has not been taken over by other scholars however, the current study will work with L1 or *first language* to refer to this phenomenon.
[7] This is also sometimes referred to as *English as an additional language* in order to avoid the impression that it has to be the second language (cf. Seargeant 2012, 192).

life, but might be restricted to official or institutional contexts. The most striking cases of this linguistic configuration can be found in former British or American colonies which, after their national independence, have kept English as an official language (Schneider 2007, 12). Maybe because of this, Götz and Schilk (2011, 80) simply equate ESL with speaking a so-called New English (see section 2.1.3); this definition leaves a lot to chance however, and does not differentiate with regard to language status and users at all.

The dimension of the official status which is so important for ESL is usually missing for people who learn English as a foreign language (EFL). The reasons for why somebody wants to learn English can vary almost indefinitely and can include personal, professional, or ideological motives (Strevens 1981, 2; Kachru 1983c, 38). English is taught in almost all countries in the world, and it is impossible to say how many people use English as a foreign language and in what contexts they do so. What can be said is that apart from the dimension of English as an official language, EFL and ESL have a lot in common, not least the fact that they are both, in most cases at least, learned at school. For those who learn English as a second language, it is also a foreign language, just with different implications. The two categories have therefore often been compared (e.g. Bongartz and Buschfeld 2011; Szmrecsanyi and Kortmann 2011), and many studies have treated them not as separate entities at all but as two ends of a continuum (see e.g. Gilquin and Granger 2011; Hundt and Vogel 2011). From the perspective of language acquisition, this indeed makes sense; it does however neglect the fact that especially in former colonial areas, "English has become a part of the cultural heritage" (Kachru and Quirk 1981, xix; see also Crystal 1985, 9). Although taught at school, it is not just a language that is learned because of its usefulness, but because it has a historical implication for the country and more often than not works as a national language, i.e. a symbol of identity, as well as an official one (Mesthrie and Bhatt 2008, 5; Seargeant 2012, 196). Due to this, these varieties of English have a certain autonomy which EFL varieties do not to that degree display: by becoming institutionalised and constituting a part of the country's historical background, English is adapted to the society and will develop characteristics that set it apart.[8]

[8] Kachru (1983b) uses the terms *institutionalised variety* and *performance variety* to emphasise this distinction.

This insight was by no means undisputed, and is closely linked to the veil of confusion that covers the term of *native speaker*. While in the cases of EFL, ESL, and L1 the concepts of the languages go hand in hand with those of the respective speakers, this is not true for *native language* and *native speaker*. As was described at the beginning of this section, the use of *English as a native language* (ENL) is fairly homogenous. Based on this, Kachru (1981, 1983c) split the varieties of English into two categories, native and non-native ones. The native ones contain those varieties spoken in the USA, Australia, New Zealand and Canada, as English was "transplanted" there, that is, large numbers of native speakers settled there to stay (Kachru 1981, 16-17). The non-native varieties include everything else, both EFL and L2 varieties (although the distinction is emphasised), and are defined to have "mainly developed in 'un-English' cultural and linguistic contexts in various parts of the world" (ibid., 15). This distinction is clear from a historical point of view and takes into account the different circumstances under which English reached foreign shores. Since then, however, things have changed; the emerging forms of ESL varieties, in contrast to EFL varieties, have developed and thus created a distance from the native varieties as they have undergone processes of "de-Anglicizing" and "de-Americanizing" (Kachru and Quirk 1981, xix). In many cases, processes of nativisation (also sometimes called *indigenisation*) can be observed, meaning that the language has developed "new linguistic features at all linguistic levels, features that would be considered deviant if used in countries where the 'native speaker' varieties of English [...] predominate" (Lowenberg 1992, 109; see also Richards and Tay 1981, 45; Kachru 1990, part 1; Mufwene 2015, 8). These varieties of English might therefore not belong to the native varieties as described above, but display a sort of "functional nativeness" (Kachru 1997, 68) due to being in frequent use and having been adapted to their environment. This raised the question of possible standards of English: If English is taught in an ESL country, should the local variety be given priority or should the teachers look to Britain or the USA for norms? And can a speaker be called a native speaker if they speak a variety of English other than British or American? In other words, which rights do these emerging varieties have, and does anyone 'own' English?

The immediate answer to this must of course be: no. "Language is an immensely democratising institution. To have learned a language is immediately to have rights in it. You may add to it, modify it, play with it, create in it, ignore bits of it, as you will" (Crystal 2003, 172). Crystal's attitude is shared by many, as by Kachru and Smith

(1985, 210), who state that "the language now belongs to those who use it as their first language, and to those who use it as an additional language, whether in its standard form or in its localized forms". It is inevitable that English changes the more it is used (cf. Strevens 1981, 1), but it is natural that this poses a problem to the *native speaker* concept. Until then, it was generally meant to refer to a person who "fully commands a language and has proper intuitions on its structural properties" (Schneider 2007, 17; see also Mesthrie and Bhatt 2008, 36; Seargeant 2012, 196). Christophersen (1988, 16) speaks of somebody who "is fully at home in the language, is confident in his use of it and is able to make judgments about usage with which other members of the language community will normally agree". With diverging varieties of English however, this agreement mentioned by Christophersen is not guaranteed, quite on the contrary; native speakers might differ greatly in what they think is acceptable use of their language (ibid.; Romaine 1992, 254). These definitions therefore bring about the problems mentioned last in the quote from Christophersen at the beginning of the section, namely those of the subjectivity and "emotional connotations" implied in the term *native speaker*. Referring to abstract ideas like "intuition" and "judgment" is unproblematic when discussing the concept in everyday life, for scientific purposes however it renders the definitions too imprecise. For the present study, *native speaker* will therefore be used in a very restricted sense and in correspondence to the *ENL* definition to refer to people who have learned English from birth.

In this sense, the categories of ENL, EFL and ESL correspond with the circles of Kachru's model. The inner circle, the norm-providing varieties, mainly consist of speakers who have English as a native language, the outer circle contains the L2 varieties which are norm-developing and are thus gaining in independence from the inner circle-varieties, and the expanding circle includes all EFL varieties, those which depend on the norms provided by the inner circle (see also Berns 2005, 86-87). In contrast to Kachru's distinction between native and non-native varieties however, the categories in his model are flexible and allow varieties to change their status. The varieties at the core can be said to have their own standards, and to use it confidently in all domains of life, and varieties from the outer circle are, by varying degrees, approaching this status, too. What is interesting is that English being a native language for most speakers ceases to be an exclusive criterion – it just happened to be so until now as the inner circle equals the original set of native varieties. With the emerging varieties gaining more independence, there could soon be a group of varieties in the inner circle in which English functions as a first or second language, and yet has developed a functional standard for the country in question. A similar trend can be observed for the movement from expanding to outer circle. In 1986, Hancock and Angogo asked "Where do learners' errors stop and legitimate features of a local English begin?" (306; cf. also Mesthrie and Bhatt 2008, 8-9), in other words, when does EFL become ESL? The major points to distinguish these two are, as was seen above, that ESL is used in a more extensive range of contexts in everyday life (often due to its official status) and that it is supposed to have historical, cultural implications that are traced back to as well as constitute national identity. The variety thus has to be accepted internally, the switch has to be made from seeking external norms to developing one's own (Strevens 1981, 2; Bolton, Graddol, and Meierkord 2011, 466).

The notion of whether or not something is acceptable of course in turn relates back to the idea of standard English. The original concept of two major standards, British English and American English, appears to have outlived its adequacy in the face of so many new varieties of the language, yet learners often request a standard (cf. Kachru 1983c, 32), and especially in EFL contexts this is still mainly sought in Britain or the USA. Furthermore, the original 'owners' of English have from time to time shown themselves hesitant to give up their privileged status among speakers of English (Kachru 1983d, 234-35; Romaine 1992, 254; Widdowson 1994, 377). This linguistic form of territoriality is curious, because the standards of British and American English have not always been unchallenged in their own countries. In Britain, the sophistication which has often been associated with speaking Received Pronunciation or 'Queen's English' has been criticised for being segregational in functioning as an "accent-bar" (Abercrombie 1951, 15; Kachru 1983a, 8), blocking the way up the social ladder for everybody underneath it. In the USA, the teaching of standard pronunciation at the cost of localised accents has been called "unintelligent" (Kenyon 1924, 3) and was feared to threaten the identity of minorities. Both standards are spoken by a mere minority of people, and are in every sense abstract sets of rules and guidelines. And yet, they are sets of rules that are fixed in dictionaries and grammars, a step which many emerging varieties of English have not yet taken.

In today's linguistic landscape, at least, the way is paved for the acceptance of emerging varieties, and they are largely met with a sense of curiosity. It has become plain that the English language is no longer defined by the original native varieties, but that it has become “a public domain” (Modiano 1999, 27). The terminology described above, the notions of ESL or L2, EFL, ENL, EIL and so on, are tools in the attempts to describe the diversity of English, although, as Strang (1970, 17) states, “language is human behaviour of immeasurable complexity. Because it is so complex we try to subdivide it for purposes of study; but every subdivision breaks down somewhere, because in practice, in actual use, language is unified”. Despite this obvious truth, the set of terms has been generally agreed upon as being helpful (Görlach 1989, 279), and will be applied in this study in the senses outlined above. It should be kept in mind however that while referring to individual speakers allows a quite clear-cut distinction, whole varieties are often generalised. Today most countries are multilingual, and can thus not be classified entirely and as a whole as an EFL variety, or an L2 society (see Schneider 2007, 13). When such terminology is used in this context, it refers to the majority of speakers in their use of English.

This holds true also for the five regions discussed in this study. The Englishes spoken in Australia, the USA and Britain are what Kachru termed native varieties, and mainly consist of native speakers of English. The other two varieties, those spoken in Kenya and Hong Kong, display a more diverse picture; both belong to the so-called New Englishes, the varieties spoken in former colonies of Great Britain and, partly, the USA. Most countries with this heritage have kept English as an official language after gaining national independence, and have made the language more or less their own. Language policies have varied greatly among the respective areas, but many of the varieties in question find themselves ESL varieties today, as English is used not only in official functions but also in cultural contexts such as literature and arts. This emancipation has been referred to by scholars as “the empire writes back” (Rushdie 1982, Ashcroft et al. 1989) or “the empire strikes back” (Romaine 1992, 47). Surely, in both Kenya and Hong Kong English has played and still plays an important part not only in everyday life, but in the shaping of the national identity. The following section will look more closely at the concept of *New Englishes* and put the varieties of Kenya and Hong Kong in the contexts of their respective histories.

2.1.3 New Englishes and the Cases of Kenya and Hong Kong

The term *New Englishes* was introduced to a larger audience by Platt, Weber, and Ho (1984). In the already vibrant market of technical terms relating to the study of the English language, it initially caused some confusion as it was used both as a synonym for *World Englishes* (see e.g. Mufwene 1994, 55) as well as a description of a subgroup of these, namely all varieties of English spoken in former colonies of the United Kingdom or, in some cases, the USA. It was the latter sense in which Platt, Weber, and Ho had originally employed it and which finally won the upper hand in linguistic discourse; yet "the obvious question of how many New Englishes there are and which countries belong to the various categories suggested cannot be answered precisely at this point. Any attempt at a comprehensive listing will require extensive discussions and unavoidably arbitrary decisions" (Schneider 2003, 237). Still, a list of certain countries, based on their colonial history, is generally agreed upon, the most striking cases being India, Singapore, The Philippines, Ghana, Hong Kong, Sierra Leone, Fiji, Cameroon, Barbados, Jamaica, Kenya, Sri Lanka, Puerto Rico, Tanzania, Trinidad and Tobago, Malaysia and Nigeria. Approaches differ however in their levels of distinction; while some studies look very closely at individual countries (e.g. Childs and Wolfram 2004 for Bahamian English; Devonish and Otelemate 2004 for Jamaican English; Asante 2012 for Ghanaian English), others summarise countries under headings such as African (Kachru and Nelson 2006) or East African English (e.g. Hancock and Angogo 1986; Haase 2004; Schmied 2006; Rooy et al. 2010)[9] or Caribbean English including, among others, Jamaica, Trinidad and Tobago and Barbados (e.g. Craig 1983; Haynes 1983; Lawton 1986; Christie 1989; Allsopp 2003).

Furthermore, opinions differ greatly on how to sort the varieties spoken in America, Australia, New Zealand, Canada and South Africa. Although Kachru's suggestion to distinguish between native and non-native varieties solved this problem, Biermeier (2008) for instance, analysing word formation in New Englishes, includes New Zealand in his study, but not the other four. Schneider (2007) on the contrary does not only include all of these varieties, but also adds to the terminological confusion by replacing the only just settled-in 'New Englishes' with 'Postcolonial Englishes'. This term, he claims, is not only more neutral but also shifts the focus to the most relevant aspect, namely the origins of these varieties in the countries' colonial and postcolonial histories and the therein embedded contact situation of English and the respective native languages.

[9] Most studies that work with this grouping of East African English are based on the ICE-component East Africa and therefore do not argue for this combination, but rely on the data provided by the ICE collection.

In order to describe this concept more clearly, Schneider, over the course of several years, formulated and refined his "Dynamic Model" of the evolution of postcolonial Englishes (e.g. 2003, 2007, 2014). According to this outline, emerging varieties, in this case of the English language, undergo five major phases: (1) that of *foundation*, in which English is established in a new area by colonial powers, (2) that of *exonormative stabilization*, in which the indigenous population and the new settlers increase their contact and external linguistic norms are accepted for the new, local English, (3), that of *nativization*, which, according to Schneider, is the central phase and is characterised by a desire for political and cultural independence as well as a rise in national pride which is taken over by the settlers' descendants, (4) that of *endonormative stabilization*, in the course of which national independence is gained and the local variety is institutionalised in dictionaries and grammars, and (5) that of *differentiatio*n, during which the new variety itself experiences a movement towards diversity through the emergence of accents and dialects.

This model certainly goes a long way in explaining the development of language varieties in colonial and postcolonial settings. It does, however, neglect certain aspects which come to mind when looking at the diversity of existing varieties of English. Not in all of the countries named above, for instance, does English function as a first language for most people. For the majority of people in most ESL-countries, it is simply a second language learned in school and serving the purpose of official language. In other countries like for instance India, there is a huge internal diversity concerning who speaks English as a first, a second or a foreign language, despite it having official status and thereby having reached, if not surpassed, stage 4 of Schneider's model. Stage 5 of the model, however, will only be reached when a variety goes beyond a purely functional purpose – as long as being the country's language of choice in matters such as politics, media and education is its major purpose, any form of differentiation would be counterproductive. For internal diversity to become a reality, it must be accompanied by the language being used in more forms of social interaction, and the more speakers take part, the faster the process will be. In this sense, stage 5 should refer not only to regional but also functional differentiation, an aspect which is being left open though.

Apart from these issues with regard to the content of the model, the introductory notion of Schneider (2007) to replace *New Englishes* with *Postcolonial Englishes* is unfortunate, as the two terms describe different phenomena.[10] Schneider's model is a purely diachronic one which could be adapted to describe varieties of other languages such as French or Spanish. It is a very helpful framework for colonial and postcolonial settings, yet in its entirety the varieties it describes equals the sum of all ESL and ENL varieties, with the sole exception of British English, and even that can be argued to have undergone a similar development (Mesthrie 2006, 381). Schneider's model can therefore be seen as a categorisation to determine the developmental status of a variety rather than its current status in relation to other existing varieties. The difference becomes clear by the fact that Schneider himself, despite criticism towards Kachru's circle-model (e.g. Schneider 2007, 12), works with Kachru's terminology, frequently referring to inner and outer circle varieties (2014, 12). Neither of the two models (or any other that have been depicted in the previous subsection) can do what the other does, and the picture of a variety is left incomplete when described along one perspective only, thus making it apparent once again that models and categorisations can only go so far when analysing language varieties.

For the purpose of terminology, therefore, the current study will work with the definition as given in the original description of 'New Englishes' in Platt, Weber, and Ho (1984). They exclude the Englishes spoken in America, Australia, New Zealand, Canada and South Africa, stating that the situation has been different in these countries, as "there has been a continuity in the use of English. People came to these areas speaking English and remained speaking English. Other immigrant groups were absorbed into the main body of English speakers" (ibid., 10). It should not be forgotten that there were, of course, native languages in these countries with which English came into contact; they were, however, less dominant than in other colonial territories, as getting into contact with the indigenous people was not a main purpose of these settlements in the first place.

To define the still large category of 'New Englishes' more exactly, Platt, Weber, and Ho (1984, 2-3) name four criteria for a variety to be called a New English, namely

[10] *New Englishes* and *Postcolonial Englishes* are not the only terms that have been suggested; Fishman, Conrad, and Rubal-Lopez (1996) talk about *Post-Imperial English*, Moag (1983) uses the term *Third World Englishes*, and Mazrui and Mazrui (1996) analyse *imperial language*. All these can be said to roughly equal the concept of *New Englishes* however, and are by far less frequently used than the two terms described above.

that it (1) has been established in the education system of the country, (2) developed in a country were English was not originally the major language, (3) is used for various functions and finally, (4) has adapted to the social and geographic environment and has therefore developed a number of characteristics of its own. It is obvious that to some degree, these factors are implemented in Schneider's model; they are more explicitly focused on the current state however, and are thus more helpful for a terminological distinction. Also, they themselves imply certain ideas that are not necessarily self-evident, especially when seen in the context of the discussion concerning World Englishes outlined above. First of all, it becomes clear that New Englishes are regional varieties rather than functional ones, as each of them is to be applicable to a number of contexts and functions. They can thus be said to include various registers; it should be noted however that the criteria given above by no means suggest that a register cannot share common ground with its equivalent in another regional variety.

Furthermore, the factors that English is used as well as taught in a country's educational institutions and that it has developed distinct linguistic features call for as well as ensure a variety's and the respective country's autonomy. For these criteria to be fulfilled, English cannot be just any language in the given country, but has to have the status of a lingua franca or, as is the case in most former British colonies, of an official language. Its institutionalisation in these countries is in line with Kachru's call for increased autonomy of the varieties, and shows that what Platt, Weber, and Ho had in mind when discussing New Englishes roughly corresponds to what Kachru would classify as outer circle and thus norm-developing varieties. Still, as helpful as these categorisations are, it should be kept in mind that no variety exactly equals another, neither in its diachronic development, nor in its developmental stage or its status within its society. The situation is different in every individual case, and the two varieties which are part of this study, the Englishes of Kenya and Hong Kong, show this quite clearly.

Kenya is, of course, by far the larger area of the two. Of the 45.5 million inhabitants, spread out over about 581,309 sq. km,[11] about 85.9% are registered as literate (UNESCO Institute for Statistics 2015; United Nations Statistics Division 2016). The two most frequently spoken languages in the country are English and Kiswahili (also

[11] Data concerning the country's exact size vary: the CIA (2016) puts it at 580,367 km², official Kenyan sources mention 581,309 km² (Kenya Presidency 2016) or 582,646 km² (Kenya National Bureau of Statistics 2014, xiii).

known as Swahili, or kisWahili); in total however, there are 67 living languages in the country and apart from English and Kiswahili, 10 others are used in educational contexts (Lewis, Simons, and Fenning 2015, 6). The linguistic situation is thus very diverse, and always has been so. The vast majority of these 67 languages (and in fact several more that are now extinct) was spoken in the area long before English arrived. Kiswahili functioned as a shared language among the people of eastern Africa as early as the 8th century, and remains one of the most widespread languages in Africa until today (Skandera 2003, 9). English first reached this part of Africa in the 16th century, but the interest of the British exceeded mere trade only in the late 19th century, during what is often called the 'scramble for Africa'. European colonial powers had discovered the continent as a new market with rich resources and huge economic potential, and claimed various parts for themselves. Finally, as a result of the Berlin Conference in 1884/1885, Africa was divided between the European nations and the area which today includes Kenya became British (see e.g. Leifer 1977, 139; Ajala 1983, 179-80; Skandera 2003, 10).

Despite the increasing number of British settlers, however, English did not spread easily among the indigenous population. When English reached this part of Africa, there was no need for a lingua franca among the people that suddenly found themselves united under the British Crown, as they already shared Kiswahili (Abdulaziz 1982, 97; Hancock and Angogo 1986, 309). Furthermore, speaking English was regarded as a prestige rather than a necessity – only very few Africans were taught English in order to work in administration, while the vast majority was intentionally kept away from the settlers' language (Abdulaziz 1991, 394-95; Mazrui and Mazrui 1996, 272; Skandera 2003, 11). It was not until Africans were sent to fight for Britain in both World Wars that this view was challenged by the indigenous population, and English began to spread, if only slightly. The final and most important reason for English to become a Kenyan official language can be found in another part of the former empire, namely India. Britain lost its most valuable colony after the Second World War in 1947, and with that had to face a weaker stance also in Africa. With more and more colonies successfully fighting for independence, Britain finally introduced English as an educational language in Kenya in order to ensure close bonds for a potential postcolonial phase (Skandera 2003, 12-13; Schneider 2007, 192). At the same time, many British settlers left Africa in the face of what might happen, thus dramatically reducing the British speakers of English while simultaneously increasing the number of African speakers. It is therefore at this time, at the end of the colonial area, that many scholars (e.g. Skandera 2003, 13; Schneider 2007, 193) see the start of Kenyan English, with a characteristic context and typical linguistic features of its own.

Despite these efforts of the British settlers, English was by no means safe. After a decade of rebellious unrest throughout the country, Kenya gained national independence in 1963 (see e.g. Hecklau 1977, 179; Fishman 1996, 6), and being fluent in English became the only language requirement for anyone wishing to be elected into parliament. This law was challenged and also changed numerous times in subsequent years; in 1974, the constitution was altered to name Kiswahili as the only parliamentary language, excluding English from this context altogether. Although this radical turn was explicitly supported by Jomo Kenyatta, the country's first president, it was changed once more only a year later; in 1975, it was declared that Kiswahili would prevail as the only language spoken in parliament, whereas English was to be the language for everything that was fixed in writing. In 1979, English was reintroduced as a language for parliamentary debate, to be used alongside Kiswahili, and both have since been required from potential candidates (Heine 1977, 266-67; Abdulaziz 1982, 99; Zuengler 1983, 114; Mazrui and Mazrui 1996, 287-89; Skandera 2003, 14-15).

Possibly due to these fast changes, there is considerable disagreement as to the exact status of English in the literature concerned with the linguistic situation in Kenya. While some scholars (e.g. Kanyoro 1991, 404; Michieka 2009, 352) state that English is an official language, others explain that it is not (e.g. Zuengler 1983, 114-15; McArthur 1998, 41). Some, as a compromise, use the term "co-official" (Abdulaziz 1991, 391; Schmied 1991a, 26; Skandera 2003, 15) or "principal language" (Ethnologue 2016). The Kenyan constitution, in its revised version of 2010, states that:

> 7. (1) The national language of the Republic is Kiswahili. (2) The official languages of the Republic are Kiswahili and English.
>
> 120. (1) The official languages of Parliament shall be Kiswahili, English and Kenyan Sign language, and the business of Parliament may be conducted in English, Kiswahili and Kenyan Sign language.
>
> 259. (2) If there is a conflict between different language versions of this Constitution, the English language version prevails.
>
> (Kenyan National Council for Law Reporting with the Authority of the Attorney General 2010)

Although the differentiation between national language and official language is thus explicit, there are no definitions provided; for the study at hand, English will therefore be referred to as an official language of Kenya.

Considering this historical framework, it is not surprising that Kenya's linguistic landscape is still so diverse today. English might be omnipresent and multifunctional in Kenya, but its way to this status was a long and winding one. Schmied (1991a, 27) claims that "calling African nations anglophone is obviously a gross exaggeration, because all of them – including the nations with a sizeable number of English mother-tongue speakers – are primarily 'afrophone'", thus relating to the more or less ambitious efforts of many African countries to promote and support local African languages, or at least save them from extinction (ibid., 24, 35; Skandera 2003, 14).

The conflicts this entails are mirrored in many studies analysing the attitude of Kenyans towards English (e.g. Sure 1989; Abdulaziz 1991) or the speaker identities that are implied in the use of English (e.g. Schmied 1989b; Schmied 1991b). Furthermore, most of the research that deals with Kenya puts it into the context of (East) African English, as was already said above, thus always taking into account the overall historical context, but often mentioning the individual countries just in passing or in the form of overviews in diagrams or tables (e.g. Hancock and Angogo 1986; Schmied 1989a; Abdulaziz 1991; Bobda 2003; Hänsel and Deuber 2013). Others look at Kenya in comparison to neighbouring countries such as Tanzania (e.g. Kanyoro 1991) or Uganda (e.g. Mazrui and Mazrui 1996), again mainly looking at historical or social factors rather than concrete linguistic evidence to explain similarities or differences. To a certain extent, combining these countries when studying their language is justified as especially Tanzania, Uganda, and Kenya are closely linked throughout the colonial period. Ever since they each gained independence however, they have been moving in very different directions (Kanyoro 1991, 402; Schneider 2007, 189), and treating them as one entity neglects the changes that have happened since 1963.

Among those studies which focus on Kenya alone the sociolinguistic perspective is dominant as well, taking into account for instance rural versus urban areas (Abdulaziz 1982; Michieka 2009) or the social contexts in which English is used (Zuengler 1983; Michieka 2005). If concrete linguistic features are analysed, the fields of phonology and grammar are the ones most extensively covered, although, as Schmied (1991a, 90) and Skandera (1999, 220-21; 2003, 61) rightly criticise, a lot of the work is impressionistic rather than based on systematically collected data. The two most thorough analyses from that perspective are Skandera's (2003) study, which looks at idioms in Kenyan English using the ICE components of Great Britain and Kenya (taken out of the East Africa component) as well as a survey, and Buregeya's (2006) paper which, with the help of a questionnaire, tries to determine grammatical features that are typical for Kenyan English. By comparison, Rooy et al. (2010), a study which takes a corpus-based, multidimensional approach, is certainly much more advanced from a methodological point of view; the authors base their research on the East Africa component of ICE however, and make no distinctions between the thus included varieties of Kenyan and Tanzanian English. In summary, it has to be said that text-based analyses on a larger scale that focus on Kenya, not East Africa, and take into account more than just one linguistic feature are non-existent.

The case of Hong Kong, albeit also based on a history heavily influenced by colonialism, differs from that of Kenya in many aspects. Hong Kong is of course much smaller than Kenya, yet it is nearly a hundred times more densely populated. On an area of 1,108 km² Hong Kong has roughly 7.3 million inhabitants, thus reaching an average of about 6,588 people per square kilometre (Kenya, by comparison, has an average of about 77 people per square kilometre) (HKCSD [Hong Kong Census and Statistics Department] 2015, 4; CIA 2016). Hong Kong, too, has two official languages, Chinese and English, both of which are used in official settings such as education and law (Hong Kong Department of Justice 2015, 3).

English, of course, is the 'younger' of the two in the area of today's Hong Kong. British trading ships landed in Macau and Canton (today called Guangzhou) in the early 17th century, and throughout the next two centuries a pidgin English developed as trade with this part of the world became ever more frequent and important for the British Empire (Bolton 2000, 267; Schneider 2007, 133). It was not free of conflict however; in Europe, the demand for Chinese goods was huge, and most countries imported large amounts of silk and porcelain, while Britain had by far the highest demand for tea. In China however, European goods were hardly needed, which resulted in a trade imbalance between the two regions. Furthermore, China accepted only silver as payment for their products, which posed a problem for Britain which had mainly traded with gold as a medium for payment. Britain had to import silver from other parts of the world in order to keep up the trade with China, a circumstance that made the business both expensive and cumbersome (Moise 1986, 28-29; Lovell 2011, 2-3; Marks 2012, 226-27).

In the early 19th century therefore, the British, in trade mainly represented by the British East India Company, began to export opium to China. The drug was sold to Chinese merchants as well as foreign traders in exchange for silver, thus reversing the flow of the precious metal while at the same time creating a drug addiction problem throughout China which grew almost out of control within few years. The Chinese reacted to this by banning the opium trade, and demanded that all stocks of opium were to be surrendered to the authorities (Fairbank 1992, 198-201; Gray 2002, 39-40; Ford 2010, 121-26). For a while, the channel to the port of Canton was closed altogether, thus blocking all trade and cutting of foreign merchants from their ships. Opium supplies were seized and destroyed in factories on land as well as on ships around the port, and British traders complained about heavy financial losses. In 1839, the conflict became violent and battles were fought throughout the next three years first at sea and later on land in what is known as the First Opium War. The British military proved superior and occupied Canton and later Shanghai. When the war ended in 1842, the Treaty of Nanjing was signed to define, among other things, the surrender of Hong Kong to British rule, thus starting its colonial history (see e.g. Moise 1986, 30; Bolton 2000, 267).

Once in possession of Hong Kong, the British took a different approach from the one employed in Kenya concerning the spread of English. Instead of making it an exquisite good provided to only a selected few, mission schools were established in Hong Kong which used English as a teaching medium. In 1911, the English-speaking University of Hong Kong was established to promote this endeavour (Bolton 2000, 267; Sweeting and Vickers 2005, 117). Of course, this by no means brought everybody into contact with English; originally, the mission schools as well as the university educated a small group of people who would then be employed in the international trade. Nevertheless, education levels began to rise, and in the 1920s and 1930s, numerous Chinese-speaking schools were opened in addition to the English-speaking mission schools (So 1992, 72; Bolton 2000, 268). In the decades to come, Hong Kong turned into a "wealthy commercial and entrepreneurial powerhouse" (Bolton 2000, 268), a factor which was boosted significantly by the China trade embargo established by the USA and the UN throughout the 1950s and 1960s (Huang 2001).

While China was isolated economically as well as politically, Hong Kong remained under British rule. Nevertheless, despite the rising worldwide tension between communist and capitalist countries culminating in the Korea War and the Cold

War, Hong Kong took some small steps towards their Chinese neighbours; in 1963, the Chinese University of Hong Kong was established (So 1992, 75), and ten years later, in 1973, while Britain saw massive protests against the Vietnam War, the colonial administration in Hong Kong published a proposal that aimed at installing Chinese as the major language in education: "We recommend that Chinese become the usual language of instruction in the lower forms of secondary schools, and that English should be studied as the second language" (quoted in Sweeting 2004, 305). There were factors which supported this trend, not least the gaining popularity of Cantonese music, films and popular culture throughout the 1960s, 70s and 80s; despite this however, the proposal met a lot of opposition from schools and parents, and the government settled on a less drastic approach in a second proposal only one year later, which suggested that every school should decide for itself (Pennington and Yue 1994, 2; Bolton 2000, 271; Chan 2002, 271). This indecisiveness proved characteristic for the colonial government throughout the years to come, as no real language policy was put in place; in 1974, at least, Chinese was finally recognised as a co-official language beside English (So 1992, 76).

By this time, whether or not English was the medium of education was still only a concern of very few. This was to change in the late 1970s, when primary and secondary school education became compulsory. This alone would have helped the spread of English like nothing before, but it was coupled with an additional factor, namely the colony's immigration rate. From about 600,000 people in 1945, Hong Kong saw an increase to about 5 million by 1981 (HKCSD 1982, 90; Bolton 2000, 268), which boosted English into hitherto unknown dimensions. In 1984, then, the colonial history of Hong Kong was equipped with a countdown: Britain and China signed the Joint Declaration, which settled the handover of Hong Kong back to China and defined it to happen in 1997 (Pennington and Yue 1994, 2). This did not reverse the spread of English which had so recently started, quite on the contrary; learning English became a priority for many in a desire to position oneself well for the future (So 1992, 86). Quite ironically, the spread of English was mingled with a discussion about the falling standards in the language proficiency (Evans 2009, 281-82). Suddenly, due to the educational reforms and the high immigration rate, many people learned and spoke English, which naturally led to a wider variation in the language itself within Hong Kong. This diversity was scowled upon and blamed for making English 'worse'. This debate, which Bolton and Lim (2000, 431) so fittingly call the "falling standards myth", was a very public one and found a lot of resonance in newspapers like the *South China Moring Post* ([SCMP] 1989), which claimed that "the decline in the standard of spoken and written English in recent decades is obvious and measurable". The blame for this lay, on the one hand, with the schools and the pupils, who were said to be taught badly and perform even worse (Bolton and Lim 2000, 434), and of course also with the government, who had failed to "stop the slide" (SCMP 1989). Although these arguments are fairly flat and were not resonated in academic literature, the SCMP made an obvious point regarding the importance of English when writing that "English is pre-eminently the language of international trade, which is, and for the foreseeable future will remain, Hongkong's *raison d'etre*" (SCMP 1986; emphasis in original).

Since then, especially with regard to the approaching independence from Britain in 1997, another aspect has come into focus, namely that of nationalism and national identity. Early studies (e.g. Pierson and Bond 1982; Bond 1985) have found that Cantonese serves not only as the major language in everyday life (e.g. Sin and Roebuck 1996, 239; Hyland 1997, 192), but also works as social glue. It has been described as "the language of the home and intimacy" (Hyland 1997, 191), while English "is an essential means to a successful career, but using it for intra-ethnic communication is regarded as being in very bad taste and an indication of severance from the Chinese community and its traditional culture" (Cheung 1985, 191). English, then, was seen as a necessity, but also as a status symbol, as being proficient in English usually went hand in hand with financial success and social prestige (Platt 1986, 384; Hyland 1997, 193). With Hong Kong being a centre for trade and business, English was thus desirable and often inevitable, despite its colonial origin and its distance from Chinese culture. Nevertheless, Cheung's assumption proved to assign English more impact than it really had; a survey conducted by Hyland (1997) of the attitude towards English among undergraduate students shows that apart from professional ambitions, the main emotion is indifference:

> While English may be rarely used for intra-group communication among Hong Kong Chinese, this study has found no evidence to confirm Cheung's (1985) assertion that its use indicates 'a severance from the Chinese community along with its traditional culture'. These undergraduates report a desire to acquire English for instrumental purposes and appear to recognise an educational and official value in English which might benefit the territory beyond 1997. On the other hand, respondents also seem resistant to the culture which the language symbolises. English bestows no particular social benefits on its users, nor do Hong Kongers aspire to embrace Western culture and its values. (207)

Pennington and Yue (1994) and Axler, Yang, and Stevens (1998) came to very similar results analysing the attitude towards English among children and young adults in secondary schools. They, too, were reported to show a high ambition to learn English, but did not feel particularly close to Western culture or regard English as a threat to their own cultural identity. A later study conducted by Evans and Green (2001) found parallel attitudes also among workers, who stated clearly that English was in use in all business-related matters, while Cantonese was the medium of choice for more informal contexts and conversations at the workplace.

Maybe it was because of this very functional and detached view of English that scholars have remarked that the variety of Hong Kong English seems to have received more attention and support from outside of Hong Kong than from within (Bolton 2000, 265). The status of English as an official language is in no doubt; however, there is an ongoing debate on the actual degree of autonomy and the uniqueness of Hong Kong English. Before the end of the colonial rule of Britain in 1997, the notion of a specific Hong Kong English was mainly rejected, not least because it was argued that if English was desired mainly for business purposes, particular Hong Kong characteristics would be of little help. Luke and Richards (1982, 55) thus come to the conclusion that English in Hong Kong follows external standards rather than develop its own, a view which is supported by Tay (1991, 327) and Johnson (1994, 182), although other scholars observed that some localised varieties of English, influenced by Cantonese, had begun to emerge (see e.g. Bolton and Kwok 1990; Pennington 1995). Almost twenty years have passed since Hong Kong's detachment from Britain however, and the decision to treat the English spoken in the area as nothing more than an outlier from Britain has since been challenged and largely rejected. In more recent literature, Hong Kong is inevitably listed as a region with a New English (e.g. Schneider 2007, 133-39; Gut 2011, 113-14), and a survey conducted by Chan in 2013 shows that tolerance for potential deviations from the British standard is increasing also within Hong Kong. The research into Hong Kong English has since shifted to the description of the variety and attempts to categorise it in the terms of Kachru's concentric circles as well as Schneider's Dynamic Model, and especially concerning empirical analyses, Hong Kong has been examined more minutely than Kenya. One major methodological advantage for the study of Hong Kong English certainly is the availability of a Hong Kong part in the ICE (see Bolt and Bolton 1996). While Kenya is combined with Tanzania in this collection, Hong Kong is represented by its own sub-corpus, and it is only very rarely referred to in the overall context of Asian Englishes (exceptions being Ansaldo 2009; Lim and Gisborne 2009; Auwera, Noël, and Wit 2012) or Chinese Englishes (as e.g. in He and Li 2009). Several linguistic aspects such as the usage of verbs (e.g. Lee and Collins 2004), pronunciation (e.g. Deterding, Wong, and Kirkpatrick 2008; Sewell and Chan 2010), syntactic characteristics (e.g. Gisborne 2002, 2009) and lexical variation (e.g. Benson 2002; Bobda 2009; Cummings and Wolf 2011) have been analysed using both quantitative and qualitative methods. Apart from this, the phenomena of code-mixing and code-switching have received a lot of attention, mainly due to the highly contextual uses of English and Cantonese (e.g. Luke 1998; Pennington 1998a, 1998b; Li 2000; Chan 2009).

Looking at the two historical frameworks of Kenya and Hong Kong, it can be said that while they share certain aspects, especially the colonial rule by the British, they have differed greatly in matters of language policy and identity. Today, English still plays an important part in both areas, but the differences are still clearly pronounced. In both cases, there are primary and secondary schools that use English as the main medium of instruction instead of Kiswahili or Cantonese respectively (for Hong Kong see Sin and Roebuck 1996, 252; Evans 2009, 291; for Kenya see Kanyoro 1991, 403; Lewis, Simons, and Fennig 2015, 33). In both cases, too, this has led to debates about language priorities. In Hong Kong, the number of parents who send their children to English-speaking schools has steadily increased until 1997 (Evans 2009, 282), and English is welcome as a means to success in a later career – yet for many children in early education it is hard to follow classes in English, which leads to the question whether it is worth putting at risk the education in order to ensure proficiency in English (Llewellyn et al. 1982, 30; Poon 2009, 23-24). Similar perspectives have emerged in Kenya; although the sense of English as a business necessity is less expressed in Kenyan society, the original British idea that English is a privilege has not yet entirely died out (Kanyoro 1991, 404; Nabea 2009, 135-36). Furthermore, it is feared that the language diversity of the country is endangered if children are not taught in their mother tongues in a school context (Begi 2014, 48).

Once again, the discussions therefore shift to the questions addressed before, questions of national identities, autonomy of regional English varieties and attitudes towards English. Today, the numbers of people who name English as their first or native language are quite small in both countries: Crystal (2003, 63) lists 150,000 for

Hong Kong and no more than 24,300 for Kenya. In addition to that, there are 2.2 million L2 speakers in Hong Kong and 2.7 million in Kenya. According to the Thematic Household Survey Report compiled by the Hong Kong government, the number of native speakers is even lower with an estimate of about 100,000. Furthermore, of people between 6 and 65 of age, roughly 17% state that they have no knowledge of English at all (HKCSD 2013, 8-18).[12] Cantonese is by far the dominant language of Hong Kong with 90.3% of the population as native speakers, and Putonghua, the standard Mandarin dialect of mainland China, follows with 3%, thereby still outnumbering English. In Kenya, 52 of the 67 active languages have more native speakers than English, although for most of them numbers of L2 speakers are not available. Even counting the 2.7 million L2 speakers of English however, the languages of Dholuo, Gikuyu and Kamba still count more native speakers by far (Lewis, Simons, and Fennig 2015, 29).

These numbers make it difficult to categorise the respective varieties in terms of a model, and they reflect once more that language diversity is a key element in both societies. Especially the linguistically diverging educational models within the two regions, but also the multilingual everyday contexts such as communication at the workplace, media discourse or family talk, question not only the categorisability of whole varieties, but also of speakers. It is therefore logical and necessary that research into these varieties looks at different facets and details of the language and the respective societies as well as keep in mind their overall structure and history. The following section will give an insight into studies that have been conducted into the field of English as a global language; since the number of works, like the field itself, is almost boundless, the studies mentioned here are not to be seen as a complete list, but as an attempt to show the most important trends in methodology and research foci.

2.1.4 Perspectives on and Approaches to English(es)

As the previous sections already suggest, the field of English as a global language is a very complex research area. It developed almost automatically out of the historical contexts brought about by colonialism in many parts of the world, and initially dealt

[12] The Thematic Household Survey Report is based on self-evaluation and should therefore be treated with some caution (HKCSD 2013, 8; see also Evans 2009, 281). People between 6 and 65 of age were asked to evaluated their proficiency in English as 'very good', 'good', 'average', 'not so good' and 'no knowledge'. 16.8% answered that they had no knowledge in written English, and 16.8% stated that they had no knowledge in spoken English.

almost exclusively with questions of politics and ideology. When English was, intentionally or accidently, exported to foreign regions by the expansion of the British and American rule, it was by no means understood that the language should settle there for longer than the settlers themselves, let alone that it should be adopted or changed. In some areas, like Kenya, English was not supposed to spread at all beyond what was absolutely necessary; in other regions, like Hong Kong, it was institutionalised more vigorously, but mainly so for economic reasons rather than linguistic ones. It was not before these colonies gained their independence from Britain or the USA that the realisation dawned that the fate of the English language was out of the former rulers' hands. The idea that English was the 'property' of England collided hard with the fact that there was no way to control it, an insight which led to fears about English varieties becoming unintelligible or the language splitting up altogether. The former colonies were also displaying very diverse language policies, and of course gained their independence at different times, and were thus impossible to coordinate, let alone streamline in matters of language planning.

The first step in this emerging research field was therefore the mere acceptance of the fact that it was indeed a research field, and a growing one at that. This presupposed resigning the idea that English would forever 'belong' to anyone, an idea which was still for a while reflected in the rather prescriptive concept of World English. Within very few years however, the idea of one World English, a "monolithic" supra-standard of the language, was first combined with and then replaced by that of World Englishes, an umbrella term encompassing all varieties of English used around the world. Once this had been – to varying degrees – accepted, the stage was set for a confusion of terminology equalled only in few other fields of study (Bolton 1999, 186; Seargeant 2010, 97). The step from World English to World Englishes was a logical one, but soon proved insufficient, and the New Englishes were set apart because of their shared historical background. This tackled only the regional perspective on language variation however, and "'Englishes' symbolizes variation in form *and* function" (Kachru 1996, 135; emphasis added). The concept of English as an International Language was introduced to ensure that everything was covered; at the same time, it was also somewhat countering the idea of World Englishes by building on the idea of an international standard based on the norms of the native varieties. Just like the regional varieties, the functional ones appeared too diverse to be a mere subcategory. *English for Specific Purposes* became the term reserved for them, and so, in most of today's literature, *World Englishes* applies to regional and *ESP* to functional varieties. Since both fields are large in themselves, the attention given to the notion of EIL has been declining steadily.

Although terminological debates sometimes still flare up or terms are used inconsistently, the focus has shifted from theoretical to empirical research. Many studies, as some of those already described in the previous section for Kenya and Hong Kong, put an emphasis on matters of sociolinguistics. The biggest topics in this branch are certainly questions of attitude towards English on the one and the power of language to shape identities on the other hand. Looking at the two exemplary cases of Kenya and Hong Kong, these issues can be said to be the logical next step after the diversity of English had been accepted, and the analyses conducted into this direction reveal highly interesting results. While some studies still put a focus on theoretical perspectives or are based on impressionistic results (e.g. Aguilar-Sánchez 2005; Leontovich 2005; Seargeant 2005), the majority of these studies draws on quantitative, rather sociological methods: Coetzee-van Rooy and Rooy (2005) for instance make use of a survey to examine how acceptable ethnic and racial categorisations like "Black South African English" are in the English spoken in South Africa. Glass (2009) combines a questionnaire with interviews to determine the status of English in Thailand and its functions as a medium of communication among Thais. Yoshikawa (2005) uses questionnaires to determine the attitudes towards World Englishes among Japanese students of English, and Moore and Bounchan (2010) question Cambodian students about their English preferences. Hiraga (2005) combines a survey with an experimental approach to determine attitudes British people display when hearing various accents, and both Richards (1998) and Lin and Detaramani (1998) conduct surveys to analyse the motivation of students in Hong Kong to learn English, and the influence extrinsic and intrinsic motivations might have on the learning process and the identification with the language.

The list of studies dealing with these dimensions of linguistics is seemingly endless, the ones mentioned above being just a small selection to display the diversity in the possible research questions. They are all located not on the level of language itself, but on a meta-level dealing with language environment. The idea of different Englishes also sparked interest in the actual linguistic properties of these varieties though, and studies started to analyse these in more detail and describe the languages. The category of World Englishes made these efforts appear rather ambitious at first;

as late as the 1990s, scholars (e.g. Kachru 1996; Crystal 1999) were uttering concerns about the sufficiency of available linguistic theories and methods to analyse this huge phenomenon. Technological advancements throughout recent years have dispersed these concerns at least partly, as the collection, storage, and processing of large datasets have become easier and cheaper. Both quantitative and qualitative approaches profited greatly from this, and thus a tremendous increase of empirical studies was triggered.

While earlier studies still relied on field work or manual analyses (e.g. Holzknecht 1989; Schmied 1991a; Bobda 2000), quantitative research received another major impulse by the compilation of the International Corpus of English (ICE). The ICE project, suggested by Greenbaum (1988), started in 1990 with the collection of comparable datasets from 12 varieties of English (Greenbaum 1990, 80). In 1996, the corpus design as well as first pilot studies were presented in a book edited by Greenbaum, and by now 14 varieties have been implemented in the collection (ICE Project 2015). These datasets, consisting of 1 million words per variety, simplified not only the description of individual varieties, but allowed comparative studies as well, and have formed the basis of many analyses since. The concrete objects of interest have varied of course; Nelson (2006) compared six varieties of English in terms of lexis, determining similarities and differences between them. Yao and Collins (2012) examined the use of the present perfect in as much as ten varieties, Nesselhauf (2009) focused on typical co-selection patterns in five varieties, including a learner corpus, and Mukherjee and Gries (2009) and Gries and Mukherjee (2010) compared four varieties, once with regard to verb construction and collostructions and once looking at preferences in terms of collocations.

Again, this is of course only a selection, but it demonstrates quite clearly one major phenomenon in variational studies. Although many datasets, and the ICE in particular, contain different text categories and therefore also hold the possibility for analyses into functional variation, rather few studies make use of this dimension. Some of the ones listed above distinguish spoken and written language in their results, or base their analysis on only one of them from the start, but only a few, like for instance Bao and Hong (2006) and Balasubramanian (2009a, 2009b), distinguish between individual text categories. Another restriction that many studies show lies in the amount of linguistic features that are being analysed. In most cases, especially in quantitative studies, the focus is clearly put on one dimension, such as a certain word class (see e.g. Bolton, Nelson, and Hung 2002, for an analysis of connectors; Mwangi 2003, for an analysis of prepositions in Kenyan English; Ersson and Shaw 2003; Sand 2004, for article usage; Mukherjee and Hoffmann 2006; Mukherjee and Schilk 2008, for analyses on verbal usage; Biewer 2011, for an analysis of modal auxiliaries; Wong 2014, for a comparison of verb-preposition constructions), individual lexical items (see e.g. Kallen 1991, for the use of *after* in Dublin English; Jeffery and Rooy 2004, for the use of *now* in British and South African English; Klerk 2005, for the use of *actually* in Xhosa-English; Balasubramanian 2009a, for the use of *also* and *too* in Indian English), or items that fulfil a very specific function (see e.g. Wong 2010, for an analysis of markers of gratitude in Hong Kong English). Qualitative studies sometimes take into account a wider part of the language, such as morphology, syntax, or phonology (e.g. contributions in Kortmann et al. 2004).

The results from these analyses are of course valuable, but they only describe very isolated aspects of language and are thus restricted in the insights they can give about a variety as a whole. One major point of interest that is not covered when researching insular linguistic features is that of the language function. English is not used in the same way and for the same matters in every variety, and the diversity of functional variation, i.e. registers, within a variety allows manifold conclusions to be drawn on the variety as well as the society behind it. The amount of situation-specific registers that can be observed in a variety and the degree of similarity or difference between these individual registers hold information as to the contexts in which the language is used and the dominance it has in the society. Furthermore, the distinctness of certain registers in comparison to others reflects in which contexts, for which topics and by which people the language in question is most frequently used. To understand the functional variation is thus elementary and can serve as an indicator for the role of English in a society and the status of the variety as such (see Mollin 2007, 172-73; Neumann and Fest 2016, 195-96).

In theory, the importance of functional variation in the framework of English as a global language has been recognised. Both Kachru (1983c, 38) and Schmied (1991a, 36) remarked that ENL and ESL varieties will display more registers than an EFL variety, but do not comment on this further, let alone analyse it empirically. The notion of ESP increased the attention given to functional variation (see e.g. Brook 1979; Escorcia 1985; Swales 1985), but so far, registers have been analysed mainly either individually (e.g. contributions in Ghadessy 1988; Holzknecht 1989; Fest 2011) or comparatively within one language (e.g. Taguchi 2002; Biber 2006a, 2006b; Guz 2009) or across two different languages in originals or translations (e.g. Haase 2013; Neumann 2013; Tereszkiewicz 2013; Diwersy, Evert, and Neumann 2014; Meinl 2014). In order to define varieties in the global context however, registers have to be compared across varieties of the same language – only then differences can be found which allow conclusions as to the developmental status and functional diversity of the varieties. For a quantitative analysis this of course implies working with a lot of data, but in recent years several studies have set out to make use of the ICE or other collections in that way (e.g. Louwerse et al. 2004; Xiao 2009; Rooy et al. 2010; Neumann 2012; Neumann and Fest 2016). Due to the amount of data needed for quantitative work in this direction, the methodology has become more sophisticated as well, relying on multidimensional analysis as developed by Biber (1988) or other multivariate techniques. Of course, a quantitative analysis is not the only possible approach to examining registers and varieties in this way; a qualitative analysis would have to be very selective concerning the data however in order to still be able to cover a number of functional and regional varieties. A quantitative approach, although prone to more generalisations, can make use of a bigger pool of available data and thus find differences "no observer would have expected […] to exist" (Schneider 2007, 87). Up to now there are, to our knowledge, no purely qualitative studies of different registers across varieties, however, qualitative elements can often be found to complement quantitative analyses by providing in-depth research of individual aspects and layers.

Unfortunately, among the studies that do take functional variation into consideration, the usage of terms such as *functional* or *register* is inconsistent. Also, it is often unclear which conditions have to be fulfilled to speak of a particular situation, in other words at what level contexts are distinguished. The following section will therefore take a look at approaches to text analysis and explain in more detail the framework which is used for this study and define its key terminology and concepts.

2.2 Analysing Texts

When looking at empirical data, as is the case in this study, we are necessarily looking at samples of the language in question. The varieties which are under examination in this thesis will be analysed with a focus on how news stories are structured and what they reveal about the language community beyond. For this purpose, the news texts which have been collected need to be examined individually as well as within their sample and with regard to their production contexts. The field of text analysis is one with a long-standing tradition; it has brought about a wide range of possible approaches and theories which emphasise different aspects of text and are thus adequate for different research questions.

The following chapter will take a look at the most influential models of text linguistics and, focusing on a functional approach to analysing texts, define the concepts and terminology relevant for the analysis. In order to put both news analysis in general and this work in particular into the context of systemic functional-related research so far, the section will end with an overview of the scientific problems and questions the framework has mainly been applied to and the methodologies with which it has been realised.

2.2.1 Approaches to Text Analysis

In the field of text linguistics, the term *text* is not always used homogeneously. Some approaches, especially from the German research tradition, use the term to refer to written language only while others, mainly from the anglophone world, include spoken language as well (Thiele 2000, 132; Neumann 2013, 10). The basis for most of these approaches was formed in the analysis of spoken language however – the earliest systematic works that are mentioned in most introductory books to text linguistics are dealing with conversations and the language used in this context. The reason for this foregrounding can be traced back to sociology and psychology rather than linguistics; in 1934 Mead defined communication as a social and interactive process with the major function of creating understanding between the discourse partners. The meaning of what is said is created in and thereby immediately linked to the situation (Mead [1934] 1967, 106-7; cf. also Heinemann and Heinemann 2002, 31), making the context essential for an analysis of language.

This view was supported by Bühler's ([1934] 2011, 175-76) work of the same year, in which he states that linguistic signs only unfurl their full meaning in relation to other signs. It was again taken up by Jakobson (1960) and Benveniste (1971), who presented models of communication similar to Bühler's, but with slight shifts in their definition and breadth of *context* (see Heinemann and Heinemann 2002, 34-35; for a detailed discussion of Jakobson's enhancement of Bühler's work, see Coseriu 1994, 76-92), and was reflected in the work by Malinowski (1935), which considers speech "meaningless" when taken out of its context. Especially the latter, originally anthropological approach, and its further development by Firth (1949, 1950) paved the way for a systemic functional framework, which will be discussed in more detail in section 2.2.2.

Across these models, communication is understood as an intentional action carried out by the speaker and, to a certain degree, the receiver. Bühler stresses this point by saying that "all concrete speech is in vital union with the rest of a person's meaningful behaviour; it is *among* actions and is *itself* an action. […] [It] proves to be directed towards a goal, towards something that is to be attained. That is exactly what a psychologist calls an action" (Bühler [1934] 2011, 61; emphasis in original). Communication can thus be said to be an example of what Weber defines as rational-purposeful action ("Zweckrationales Handeln", Weber [1921] 1976, 32-33; cf. Steiner 1991, 87; Heinemann and Heinemann 2002, 38), an assumption which almost inevitably leads to the idea of utterances as acts.[13] The first to formulate this concept was Austin ([1962] 1973), who examined performative speech acts and the effects on reality caused by them. From the general insight that language can indeed change the world just like other actions, Austin extracts the definition of three speech acts, the locutionary, illocutionary and perlocutionary one, which to him represent different sides of the same utterance (Austin [1962] 1973, 94-101; see also Heinemann and Heinemann 2002, 40-41). Furthermore, he defines basic conditions that have to be fulfilled for the communication to be "felicitous", i.e. successful (Austin [1962] 1973, 14-15).

This multidimensional picture is partly taken up by Searle (1969), who changes Austin's model slightly by neglecting the concept of the perlocutionary act and instead assigning a more key role to illocutionary acts (Heinemann and Heinemann 2002, 41), thereby shifting the focus to these aspects of an utterance which most clearly involve both speaker and receiver and signalling that the recipient is actively involved in the communication process. This part is emphasised even more by Grice (1979b), who

---

[13] Werlich, with a clearer focus on linguistics than was intended by Weber, distinguishes social action from communicative action and defines the latter as being characterised by the intention of interaction with others. Within this category, he separates verbal and non-verbal actions, both of which can appear isolated or in combination and create instances of text (Werlich 1975, 13). Although Werlich does not refer to Weber, it is worth noting that communicative action in his sense can be regarded as a subcategory of rational-purposeful action (cf. Gülich and Raible 1977, 23).

focuses his work on indirect speech acts, intentions that are not explicitly uttered in a conversation but are implied in what is being said. The recipient in this case has the responsibility to infer what the communication partner might intend, and a lack of motivation or ability on the recipient's side might lead to misunderstandings and failures of the communication. In order to do this, the recipient must in turn be able to rely on the speaker to make his intentions sufficiently, if not explicitly, clear, which is possible because we have standard ways of expressing certain intentions and wishes (Grice 1979a, 14; cf. Heinemann and Heinemann 2002, 42-44). The recipient therefore gathers the knowledge they need not only directly from the speaker, but from the context of the communication in general. Although the maxims presented by Grice certainly draw on the foundation laid by Austin in his felicity conditions, the receivers' part is acknowledged as more active in Grice's understanding and entitles and obligates them to more than just reacting.

These models and approaches, as was said above, are all aimed at spoken language, and work with the assumption that speech is a more or less purposeful action with an intention behind it. Although scholars attribute varying degrees of importance to the recipients, their presence is always taken for granted at least as the party to whom the intention is addressed and who is expected to react to it in some way. Speech as action, if successful, triggers a reaction (Gülich and Raible 1977, 30), and for this mechanism to work a dialogic structure is necessary. When looking at written texts, the constellation is much more obscure; arguably every text is intended for some recipient, but writer and reader are usually separated in space and time. A reaction can therefore only appear with delay, and as soon as texts are addressed to a larger, often anonymous audience, as is the case with for instance books, newspapers, brochures, but also signs or manuals, a reaction from the readers is close to impossible, at least in a way that would be perceivable for the author. Early approaches to written text therefore concentrate less on the idea of speech (in this case writing) as an action but focus instead on the structures of the texts and the way they transport their meaning and influence their readers despite spatial and temporal discrepancies.

The first approaches to this endeavour analysed texts on the basis of grammar. Isenberg (1968, 1971) draws on the principles of generative grammar, applying the rules of sentence structures to sentences in a given text and roofing them with a similar concept of text structure rules (Heinemann and Heinemann 2002, 65-66). This, in his sense, makes it possible not only to describe an already existing text but also to formulate guidelines how to produce an ideal text. The approach emphasises the structure of writing, but neglects the semantic level and fails to take into account the actual content of a text. A slightly broader, yet still grammatical approach is suggested in Harweg's (1968) work, who, while also relying on the basic idea of sentence structures, sees the links between the sentences not reflected in a superimposed text structure but in cohesive devices. He identifies the length of a text by the length of chains of pronominal constituents; as long as there are references to a certain item, usually in the form of pronouns, the writing can still be said to be a meaningful unit and therefore be called a text (Harweg 1968, 148; see also Mehler and Köhler 2007, 7). For this analysis, the content has to be taken into account at least to the degree of identifying intratextual references and their origins, however, the structure remains the focus of the approach.

A shift in focus to a more semantic analysis of texts was triggered by Greimas ([1966] 1983). Making use of Harweg's concept of chains, Greimas loosens the idea from its purely grammatical level and instead identifies semantic chains, that is, sequences of coherent strings of content, which he terms isotopies. These cannot be described on a structural level alone but require a close examination of the text's content beyond the level of individual word classes or grammatical categories, yet it still assumes a text working in a linear way – any bit of content that does not immediately fit into an isotopy will either be left unconsidered or necessitates the definition of a new semantic chain within the text (see Heinemann and Heinemann 2002, 72-73). This leads to chains of various lengths and complexities, which in itself allows insights into the text and helps an analysis; however, if chains become too detailed and numerous, the approach runs the risk of losing its applicability and generality and turns into an unsystematic analysis. Furthermore, the model emphasises the semantic chains and their coherence throughout the text, but does not give much thought to the way these chains interact and might be foregrounded or eclipsed depending on the way they are employed in the text.

The idea that specific formulations and syntactic structures can have an influence on the meaning of a text was introduced by Daneš (1970, 1976), who in his works relies on the functional sentence perspective as formulated for instance by Mathesius (1929) to analyse the structure of sentences in a text and determine which information is emphasised as the theme by the syntax. By defining this for sequential sentences, a thematic progression can be identified beyond sentence level. The theme in this context can be said to act as a fixed point which the reader is familiar with, while the rest, the rheme, constitutes new information on the theme (Daneš 1970, 73; see also Heinemann and Heinemann 2002, 70-72; Schubert 2008, 75).

The notion of a text's theme has since been taken up by several scholars. Van Dijk (1980, 1997) defines it as part of a text's superstructure, which influences the micropropositions (sentences) and macrostructure (the text's meaning) (see Heinemann and Heinemann 2002, 77-78). Brinker ([1985] 2005, 57) on the other hand regards the theme as the direct partner of the text's function, the two major elements which shape the text and constantly condition each other. Especially the latter element, the text's function, gives the analysis of texts a more pragmatic dimension than can be found in many other approaches and relates back to the idea of speech as action. From Brinker's perspective, every text displays one major function which is dominant. This function does not necessarily have to correspond with the speaker's main (and potentially hidden) intention, but is defined solely on the basis of which intention is shown in the text (ibid., 101). Like Grice, Brinker thus assigns the responsibility for successful communication to both speaker and receiver: it is the speaker's duty to include in their text everything the receiver needs to infer the intention, while the receiver actively has to make use of this information to understand the message. Not everything has to be included directly in the speech or writing however; Brinker (ibid., 106) also emphasises the importance of contextual indicators as means of creating and deriving meaning.

This inclusion of the context, in which the text appears, opened a new dimension of text analysis. So far, most approaches had focused on grammatical or semantic elements, and context was mainly discussed either with regard to spoken language or, as a complete contrast, as an almost exclusive criterion for text analysis, pushing purely linguistic concerns to the very edge of the picture (e.g. Ehlich 1980, 1986; Hartung 1987). A notable exception to this is the approach of Critical Discourse Analysis (CDA), which regards discourse as an "instrument to gain insight into societal problems" (Renkema 2004, 282). The main aspect of study in CDA is the relation between language and power, and the way the former can be used in order to influence readers or listeners to view the world in a particular way. Reality is regarded as being constructed by language, and any perception as well as production of text is affected by the context at hand as well as by the social roles we and others assume (Fairclough

1989, 1; 1995, 132; Locke 2004, 2). Any analysis of language in an isolated state cannot be defined with regard to its social values and functions however, and the approach is thus by necessity interdisciplinary and draws on concepts from sociology and philosophy (Wodak and Meyer 2009, 2). The context of language is considered elementary, and "discourse is studied [...] with reference not to differences between language systems, but to differences in language use within one language" (Renkema 2004, 282). CDA, with its clear focus on the power structures between a text and its producer on the one and the receiver on the other hand does not immediately offer itself to an analysis of language in use in a broader context; however, the model marks the importance of separating the abstract concept of *discourse* from that of *text*.

In anglophone research, *text linguistics* is often equated with *discourse analysis*, and many scholars use the key terms *text* and *discourse* interchangeably (e.g. Stubbs 1983; Chafe 1992; cf. also Stubbs 1996; Thiele 2000, 132; Widdowson 2004, 2-5). Others do make a distinction, but often the definitions in use are not consistent. As Hoey (1991, 197) points out: "It is as if some basic differentiation is felt to exist that people cannot quite agree on but cannot leave alone." A very straightforward suggestion is made by Esser (2009, 2-3), who draws on dictionary definitions of the two terms and remarks that while *text* refers to a "meaningful unit" of written or spoken language, *discourse* "is a (medium-independent) topic-bound exchange of views in which more than one party is engaged". Although these two definitions do not seem to complement or exclude each other and are therefore of limited help to clearly distinguish the concepts, they imply that *discourse* is something bigger than *text*, a larger entity with less distinct points of beginning and ending. Widdowson (2004) also structures the terms hierarchically, although his focus is more clearly on the function and implication of language and closer to the terminological usage in CDA: in his sense, *discourse* can be said to be language in the setting of its context and social intention, while *text* is the representation of discourse in physical, that is spoken or written, form (8). In the study at hand, the terms will be used in such a hierarchical sense as well; the individual newspaper articles will be regarded as texts, that is as meaningful units with a clear beginning and ending. The term will thus be used in a countable form, while *discourse* will remain a non-count noun in order to retain its abstract and medium-independent character (see Esser 2009, 3) and denote that the texts representing the particular field of news discourse are indeed related in their origin as texts from news.

Despite a distinction of the terms on a theoretical level, it is almost impossible to separate them in an actual analysis. Although some approaches take context, which is an essential part of discourse, into account more than others, linguistic characteristics of language are still analysed on text level, and it has to be acknowledged that despite being part of the same discourse, texts can be quite distinct. *Discourse* as such refers primarily to a relation of texts on a macro-thematic level, in the case of this work for instance that all texts are newspaper items. The topics the individual texts treat, the target groups they address and the linguistic particularities they display can vary greatly however, which shows that the gap between the concepts of *text* and *discourse* is too big to allow a concrete analysis. The immediate context is set above the purely textual level, which can be seen as a realisation and reflection of it; it is also set below discourse level however, although it is certainly conditioned by it. The approach of CDA, besides its above-mentioned restrictions due to the ideological focus on power relations, does not specify the context in a systematic way, relying mainly on abstract features when analysing the socio-political dimensions in and around language. Yet for the current study, the context is essential and needs to be analysable comparatively by means of its linguistic reflection on a textual level. An approach that offers this possibility more readily is the register framework from functional linguistics, which defines the context of situation along certain parameters and thus ensures comparability across texts. Furthermore, it is by definition designed to look at language in use and is thus suitable to map the differences between instances of language in an empirical analysis.

2.2.2 Text in Context – A Functional Perspective

The notion of register is based on the assumption that language changes depending on the situation in which it is used. It was introduced by Halliday as a development of ideas first presented by Malinowski (1935) in his work *Coral Gardens and their Magic* (see Halliday 2006, 444), in which he emphasises the importance of context for the interpretation of language, saying that “speech is meaningless without the context of the activity in which it is enveloped” (Malinowski 1935, 8). Being an anthropologist himself, Malinowski puts the focus of his work on cultural and sociological aspects, and his interest in language is always in connection to anthropological questions; he therefore distinguishes between two different contexts that are relevant for language, namely the context of situation and the context of culture. While the former refers to the immediate setting of the communication, the latter is to be seen as a much more extensive concept, representing “what speech achieves in a […] culture” as well as the “material equipment, the activities, interests, moral and aesthetic values with which the words are correlated” (Malinowski 1935, 22). Especially from a social anthropological perspective, the context of culture is of course of great interest, and, although the concepts determining it are highly abstract, it is a relevant notion for analyses in the contexts of for instance sociolinguistics or translation studies. For a more concrete analysis of particular instances of language however, the context of situation is more relevant and was as such taken up in the linguistic works by Firth.

Firth himself attributes his interest in Malinowski’s ideas to the unusual focus of his studies: “A great deal of abstract sociology is of doubtful value to the linguist because of the sociological neglect of persons and people, consequently of language also. Not so Malinowski, who gave us an ethnographer’s theory of language. He was a close student of persons and people” (Firth 1949, 399). To Firth, language is thus inseparably linked to the people using it, and despite his criticism of the abstract theories presented in sociology, his perspective, like that of Bühler and Mead, regards language as an action, although he puts less emphasis on the purpose and considers the consequences of an utterance instead. Staying true to his critical view on abstract theory however, Firth’s praise of the ethnographic nature of Malinowski’s work, which describes his field observations on cultural and societal life on the Trobriand Islands in Papua New Guinea, is logical, and he transfers the idea of a systematic framework from Malinowski’s large-scale study to the narrower field of linguistics. Relying on the concept of the context of situation as the determinant factor of language use, Firth states that it “is best used as a suitable schematic construct to apply to language events” (Firth 1950, 43), thereby positioning it as a model for empirical analyses. According to him, the context of situation should consider three aspects (ibid.; see also Fest 2011):

A. The relevant features of participants: persons, personalities.
   (i) The verbal action of the participants.
   (ii) The non-verbal action of the participants.
B. The relevant objects.
C. The effect of the verbal action.

Along these parameters language activities can be categorised and grouped; in order to do so, however, they have to be analysed linguistically on a less abstract level, requiring a “hierarchy of techniques” (Firth 1950, 44).

The basic assumption that language cannot be described properly without taking into consideration its context was taken over and developed further by Halliday. Stating that "language comes to life only when functioning in some environment" (Halliday 1974, 28), Halliday reanimates the idea of a purposeful function in language the understanding and mastering of which he considers necessary for every member of a speech community: "Being 'appropriate to the situation' is not some optional extra in language; it is an essential element in the ability to mean" (ibid., 35; see also Halliday 1978, 61). When communicating with others, we adapt our language to the situation at hand and form expectations about the discourse on that basis (Halliday 1974, 53; Fest 2011, 10). Being familiar with the situation type helps in the process, whereas a misconfiguration of our language can lead to unsuccessful communication and misunderstandings. The context of situation is relevant therefore because it conditions the language in use and vice versa, and it is this interplay which Halliday (1961, 269) spotlights: "Context is an interlevel […] since it relates language to something that is not language; it is an interlevel because it is not with the non-language activity itself that linguistics is concerned but with the relation of this to language form" (cf. also Butt and Wegener 2007, 593).

The idea that context influences language to such a degree that conventionalised situations can be recognised by speakers, causing them to adapt their language and form expectations about the communicative action, implies that these situations will be reflected linguistically. Furthermore, it presupposes that situations recur and that every society thus has a repertoire of settings which require, or at least trigger, different usages of language. To describe these functional varieties of language, based on when, where and how the language is used, Halliday introduces the term *register* (e.g. Halliday, McIntosh and Strevens 1964, 87; Halliday 1974, 32). Of particular importance in this framework is the notion suggested by Firth that although our lives consist of countless different situations, they can be grouped: "Looking at how people actually use language in daily life, we find that the apparently infinite number of different possible situations represents in reality a very much smaller number of general *types* of situation" (Halliday 1974, 29; emphasis in original). These types then constitute the registers a language contains, and the sum of all the registers is taken to equal the language activity of the speech community in total (Halliday, McIntosh and Strevens 1964, 89).

In order to allow studies to make the transition from context to language, a systematic conceptualisation as mentioned by Firth is indispensable. Based on Firth's categorisation, Halliday introduces a slightly altered tripartite model of contextual parameters, defining his variables as *field of discourse* (representing the ongoing activity), *tenor of discourse* (representing the participants and their relationships) and *mode of discourse* (representing the part language plays in the interaction) (Halliday 1975, 130-31). His categories are broader than Firth's, requiring a set of subdimensions, such as hierarchical structures or power relations as part of the tenor of discourse, to further distil their contents. Due to the increased width of the parameters however, Halliday's model is also capable of describing the context of situation more fully so that between them, the three parameters, through their various constellations, can describe every register of a language by constituting its contextual configuration (Halliday 1974, 50; Halliday and Hasan 1985, 55-56; Halliday and Matthiessen 1999, 321; Butt and Wegener 2007, 590). This configuration can be found to be reflected in the language itself: "The linguistic features which are typically associated with a configuration of situational features – with particular values of the field, mode and tenor – constitute a REGISTER" (Halliday and Hasan 1976, 22; emphasis in original).

Although this representation of context in the language is clear and logical, the concept is less definite about the question whether registers are then to be defined language-externally or internally. In order to group texts into registers, an analysis can either start by examining the contextual configuration and determine similarities on this level, or by investigating the linguistic properties of texts and screen for likenesses at this end. Over the years, researchers have adopted different perspectives on that matter and thus created a set of terminology which is not always used consistently. Especially the two concepts of *register* and *genre* are often named together while frequently adopting different meanings, both within and outside of the field of systemic functional linguistics.

The term *genre* is of course an important concept not only in linguistics but also in literary theory, where it is used with "relative stability to describe formal features of a text" (Kress 1993, 31). In linguistic theory, its descriptive frame is less clear, but many definitions that can be found attempt to include the cultural aspect implied in the term's classical use in literary studies. Couture (1986, 82) for instance claims that "genre can only be realized in completed texts or texts that can be projected as complete, for a genre does more than specify kinds of codes extant in a group of related texts; it specifies conditions for beginning, continuing, and ending a text". This view, which tends towards the literary idea of *genre*, is repeated by Biber and Conrad (2009, 2), who add that the parameters for defining genres are mainly based on text structure and can be traced back to social conventions rather than functional utility. In their view, genre and register analyses represent two perspectives on the same text, the former dealing with complete texts, the latter with intratextual aspects. While this complies with the perspective offered by Kress and Hodge (1979, 33) that genres reflect "the structural features of the specific social occasion in which the text has been produced", Biber, in earlier works, does not focus his concept of genre on structural elements, but states that "genres are defined and distinguished on the basis of systematic nonlinguistic criteria" (Biber 1989, 39).

This idea of genre is put forward in other studies as well; Lee (2001, 38), drawing on Swales (1990), suggests to regard genres "simply as categories chosen on the basis of fairly easily definable external parameters". He goes on to say that "genres also have the property of being recognised as having a certain legitimacy as groupings of texts within a speech community (or by sub-groups within a speech community, in the case of specialised genres)". This legitimacy implies a cultural and ideological nature of *genre*, and indeed can be seen to reflect the original idea of the context of culture as voiced by Malinowski. Halliday and Hasan (1985, 7), discussing the distinction between the context of situation and the context of culture, come to the conclusion that while for the analysis of some instances of language, especially highly pragmatic ones, the context of situation suffices, other, less pragmatic texts need to be embedded in their broader context of culture in order to understand their full meaning. This is due to the fact that situations are culture-specific, and a certain text which might be legitimate in one culture can be non-existent in another, simply because the situation itself is not part of that culture's semiotic potential (ibid., 100). Eggins and Martin (1997, 243) therefore suggest that "analysis at this level [of genre] has concentrated on making explicit just which combinations of field, tenor and mode variables a culture enables, and how these are mapped out as staged, goal-oriented social processes" (see also Martin 1984, 25; Eggins 1994, 33-34; Bateman 2006, 178; Nunan 2007, 209).

While the cultural aspect of genre is thus, in varying detail, repeated in most definitions of the term, its relation to the concept of register is less agreed upon. In the approach adopted by for instance Martin and Eggins, genre is to be placed on a level above register, it constitutes "register plus communicative purpose" (Thompson 2004, 42). Eggins (1994, 26), too, defines genre mainly on the basis of the overall purpose of the discourse, and refers back to Martin (1984). Martin and Rose (2008, 6) meanwhile speak of goal-orientation rather than purpose, justifying this term by stating that "we feel frustrated if we don't accomplish the final steps", thereby again emphasising that genre is a categorisation of a whole text, not of text passages. Although whether a systematic distinction between the terms of *goal-orientation* and *purpose* is intended or not remains unclear, Martin connects the former more clearly with the level of register, and thus the immediate situation, while the latter is applied in a more general framework: "The register variables field, tenor and mode can then be interpreted as working together to achieve a text's goals, where goals are defined in terms of systems of social processes at the level of genre […]; genres are social processes, and their purpose is being interpreted here in social […] terms" (Martin 1992, 503).

In partial contrast to this, Halliday (1978, 134, 145), while acknowledging the purposefulness of genres, puts the main focus on the structure of the text and thus connects genre with the textual metafunction and the register parameter of mode. It is not treated as exclusive to mode however, but allows room for connections to the variables of tenor and field as well. Despite stating that "the generic structure is […] language as the projection of a higher-level semiotic structure" (ibid., 134), Halliday thus paints the picture of a close interdependency between the planes of genre and register. This notion is increased in the later work by Halliday and Hasan (1985), in which the elements of texts are divided into obligatory and optional elements; a genre, as represented by its generic structure potential (GSP), "is the total range of optional and obligatory elements and their order in such a way that we exhaust the possibility of text structure for every text that can be appropriate to (a certain) C(ontextual) C(onfiguration)" (ibid., 64). The GSP is thereby made the main instrument of distinguishing genres: "texts belonging to the same genre can vary in their structure; the one respect in which they cannot vary without consequence to their genre-allocation is the obligatory elements and dispositions of the GSP" (ibid., 108). The GSP can be regarded as conditioning the register in use, however, it also depends on the contextual configuration, which represents the register. The relation between the two concepts is thus not that of one level being superimposed on the other, but a circular one derived from the parallel and interdependent developments of culture and language themselves (Halliday and Hasan 1985, 104). While genre can, to put it very bluntly, be connected to the context of culture, register is positioned in the context of situation; and while genre refers to the structure of complete texts and, in Hasan's conceptualisation, to the set consisting of every *potential* text in a given cultural framework, the register represents the immediate environment along the more concrete variables of field, mode, and tenor.

In theoretical terms, the two concepts are thus, despite all confusion, not treated as synonyms. Especially across corpus-based studies however *genre* and *register* are often used interchangeably for categories within a corpus that are, of course, defined externally, as the compilation of a corpus does not allow any other approach; a linguistic analysis is only undertaken after the compilation, and the text categories a corpus is to contain are therefore predefined along non-linguistic criteria. In the naming of these categories both terms, *genre* (as in e.g. Biber 1986; Xiao and McEnery 2005; Baker 2006) and *register* (e.g. Biber 1993b, 1995; Kim and Biber 1994; Barbieri 2005; Liu 2008), are frequently encountered, and several studies switch between the terms, using them as synonyms (Biber 1993a, 1995; Gries, Newman and Shaoul 2011; Baker 2010b). Especially Biber, who has often used both terms "with no implied theoretical distinction" (Biber, Connor, and Upton 2007, 9), has criticised a predefinition of texts in general, by pointing out that this can lead to a categorisation which influences the interpretation of the results, not leaving room for any conclusions that might fall outside what has been predefined (see e.g. Biber 1986, 386-87; 1989, 5; 1990, 261-62). In his approach, Biber does not rely on an a priori sorting of texts but rather processes all texts alike to let patterns emerge due to linguistic similarity (see e.g. Biber 1985, 1988, 1995; contributions in Conrad and Biber 2001). It is certainly true that a sorting before the actual analysis can, if not taken into consideration, influence the interpretation of the results greatly; however, leaving aside any knowledge of categories, especially if they have been assigned by parties outside the research, can lead to a neglect of certain aspects of the texts' production circumstances which might not be reflected in the language, but might still bear on the interpretation. Furthermore, although Biber's method includes a large amount of linguistic features to be tested in the texts, they are not accounted for in a systematic framework of the context, but are analysed in connection to each other. This inductive approach no doubt produces valuable and highly sophisticated results and insights; yet for the present study, which is to include a sociolinguistic as well as media-linguistic perspective, it will be of more significance to be able to trace similarities and discrepancies between the texts back to precise features of the context as represented by field, mode and tenor and their respective subdimensions.

Due to these research goals and theoretical foundations, the corpus used in the study at hand has been enriched with mark-up to allow pre-sorting of the newspaper articles on various grounds (see section 4.2 for a closer description of the dataset). From a terminological point of view, the categorisation of the articles based on their newspaper area is the most critical one. As was shown above, the discussion of the exact frameworks of genre and register is a very active one, yet there is still a lot of ambiguity in the use of the terms and the respective implications (Nunan 2008, 58; Fest 2011, 11). It is not the aim of this study to fully reproduce, let alone conclude this discussion. The notion of genre, alluding to the cultural potential of a language and its speech community, will be of little consequence for this work, but where it is used it will be applied in the sense of Halliday and Hasan (1985), that is as a framework for text potential reflecting cultural dimensions. Register, on the other hand, is a crucial concept for this work, as it represents language in its different uses. Resuming the opening assumption that registers will form more clearly when a situation recurs and thus becomes conventionalised, exactly this formation of different registers and their degree of distinction will be used to analyse the developmental state of the varieties included in this study. To achieve results as concise as possible, the linguistic features to be analysed will not, as in Biber's case, be queried inductively, but with clear reference to the contextual determinants of field, mode and tenor to allow conclusions as to the nature of differences between texts (see Halliday and Hasan 1976, 21). However, the individual categories as defined by the newspapers – sports, economy, politics, features and hard news – will not be treated as registers necessarily given. Despite their being allocated to similar topics by the newspapers themselves (cf. section 4.2.2), the variation within these categories with regard to their situational context and thereby their respective register features will be analysed as well as their potential similarity to texts from other categories. In order to avoid the use of a term which already bears so much meaning in linguistics, the categories will be referred to simply as *domains*.

2.2.3 *Trends and Foci in Register Analysis*

Despite the terminological disagreement and confusion, registers (or genres, depending on who is reporting) have not only been considered from a theoretical point of view. Because of their reflective character of social settings and human behaviour and relationships, they have attracted a lot of attention as bases for empirical research as well. In this sense, registers can be – and have been – analysed from various perspectives; they can be described individually or in comparison to other registers, diachronically or synchronically, and within one language, across varieties of it, or even contrastively across two or more languages. Furthermore, they can be analysed with regard to particular domains or topics, such as for instance academic language, legal language, media language or literature. Fields like these offer room for numerous registers, and have been analysed in terms of internal register variation, often with a focus on spoken and written components. Usually, studies merge several of these perspectives, which makes it very hard to simply categorise what has been done so far; in order to give the reader an overview however, this section will show the most dominant trends and methods with the help of a selection of respective literature from the field.

The first research direction that becomes apparent in register analysis is a historical one (cf. Biber and Conrad 2009, 286-88; Schubert 2016, 5). By analysing language from a specific period of English, studies aim at drawing conclusions about the relevance of the respective areas and societies. In this context, individual registers have been analysed separately, for instance letters from Early Modern (Fitzmaurice 2002; Nevala 2004) and Late Middle English (Bergs 2004), pamphlets (Claridge 2005) and courtroom discourse (Archer 2006) from Early Modern English, medical charms in Middle English (Alonso-Almeida 2008) or 18th century newspaper headlines (Studer 2003). Furthermore, whole areas have been examined with regard to their internal register variation, such as fiction from the 19th century (Egbert 2012) or scientific language in the Middle Ages (García 2004). Still other studies (e.g. Geisler 2002; Culpeper and Kytö 2010) put their focus on the differences between spoken and written registers rather than on individual fields.

Apart from these historical studies with a synchronic approach, several diachronic analyses have been conducted. Biber and Finegan (2001) look at the distinction of spoken and written register variation and trace their developments from the 17th to the 20th century, and Warner (2005) compares the use of the supportive form of *do* in two periods of time from the 16th and 17th century, evaluating its potential function as a register marker. Individual registers, too, drew some scholarly attention, with scientific registers being a very dominant choice: both Atkinson (1999) and Gross, Harmon, and Reidy (2002) look at the development of scientific prose over a span of 300 years, while Kopple (1998) and Salager-Meyer (1999) narrow their analysis down to research articles from physics and medicine respectively.

In present-day English, registers have also been analysed frequently and with similar approaches, namely individually, comparatively or within particular fields. As Biber and Conrad (2009, 271-95) and Schubert (2016, 6) point out, especially scientific language has been a focus with regard to present-day English, too; analyses have been conducted into the register of research articles, often with a focus on a particular discipline such as biology (Conrad 1996; Hyland 1998), biochemistry (Kanoksilapatham 2005), medicine (Ferguson 2001; Fryer 2013), geology (Dressen 2003), art history (Tucker 2003) or linguistics (Ruiying and Allison 2003, 2004; Ozturk 2007), but also across fields (e.g. Schutz 2013). Furthermore, specific parts of research articles have been examined, most prominently their introductions (Bhatia 1997; Gledhill 2000; Samraj 2002), discussion sections (Holmes 1997) and respective abstracts (Stotesbury 2003). Similar foci can be found regarding PhD theses; while Charles (2003, 2006a, 2006b) contrasts various linguistic features in theses from material science and political science, Thompson (2005) looks at intertextual references in works from agricultural botany. Hyland (2004) and Bunton (2005) again focus on particular parts of the text, the former on acknowledgements in PhD theses, the latter at their conclusion chapters. Other studies into particular registers of academic discourse include analyses of lectures (Csomay 2002; Thompson 2003), grant proposals (Connor and Mauranen 1999) and note-taking of students (Janda 1985). Apart from these very specific works, there are several studies which do not concentrate on one register, but rather look at academic language in general and compare different settings (e.g. Swales 1990; Bhatia 1993; Biber 2006a, 2006b; Gotti 2012).

Of course, other areas apart from academic language have been analysed as well. Two other well-represented fields are corporate discourse and new forms of communication due to technological progress. Concerning the former, especially intercultural aspects are of interest; Bargiela-Chiappini and Harris (1997) for instance analyse the discourse of corporate meetings in Italy and the UK, and Koester (2006) compares office conversations from the UK and the USA. Other aspects include letters of application (Henry and Roseberry 2001) and recommendation (Precht 1998) as well as corporate reports on social issues and responsibilities (Fuoli 2013; Fuoli and Paradis 2014), to name but a few. The second field mentioned above, new forms of electronic communication, often relates to professional discourse for instance in the form of business e-mails (Gimenez 2000; Jensen 2009) or computer conferences (Yates 1996), but is of course also present in personal contexts. While Crystal (2001) analyses and compares various online registers, other studies focus more clearly on the description of individual ones such as weblogs (Herring and Paolillo 2006), dating chats (del-Teso-Craviotto 2006) and instant messaging via computer (Fox et al. 2007) or phone (Thurlow 2003).

Throughout all these fields and numerous more in which registers have been analysed, a variation of methods has been applied. A vast majority of studies uses corpus-based approaches, which is particularly helpful when comparing several languages, varieties or registers and thus working with large amounts of data. On this basis, Vassileva (2001) compares academic writing in English and Bulgarian, Peterlin (2005) looks at research articles in English and Slovene, and Barron (2012) examines public information messages in German and Irish English. Apart from this, multilingual corpora are particularly valuable for translation studies, although the collections for studies in this field are usually more complex and contain subcorpora for original texts and translations to and from the respective languages. Teich (2003) makes use of such a collection from the register of popular scientific writing in German and English, and Neumann (2013) draws on a corpus of eight registers from these two languages.

In contrastive studies like these, register shifts are often observed when a text is being translated from one language to another (see also Moe 2010). Every language is adapted to its particular society, and therefore forms its own register constellation based on the speakers' needs and behaviour. The same holds true for regional varieties of one language, which makes register analysis a useful tool for this perspective as well. In section 2.1.4 many studies into varieties of English were mentioned already, and it was observed that many rely on descriptions of individual linguistic features rather than whole registers or even a comparison of several ones. Even rarer are works that take into account different varieties as well as registers. The compilation of bigger corpora and more fine-grained annotation systems make queries more effective and thereby such large-scale comparisons easier; studies which have made use of this are for instance Xiao's (2009) multidimensional analysis of five varieties and twelve registers, Neumann and Fest's (2016) comparison of medium in five registers and six varieties, and Neumann's (2012) analysis of field, mode, and tenor of discourse, also in five registers and six varieties. All three studies draw on data from the ICE, which distinguishes regional and functional variation in its design. Furthermore, a study by Hardy and Friginal (2012) compares two online registers in Filipino and American English, using data from a collection of blogs and opinion columns.

Although the results from these studies allow important insights into the links between regional and functional variation, they often do not discuss in more detail the cultural backgrounds of the respective varieties or the specific contexts of the registers included; collections like the ICE make it easy to analyse and compare registers without considering for instance their production circumstances or their history in the respective regions. This tendency can also be observed in studies concentrating on, or including, media language. Mass media texts have received quite a lot of attention in terms of register analysis (cf. Schubert 2016), often with a focus on particular domains such as advertising (Bruthiaux 1994, 1996, 2000; McQuarrie and Mick 1996), editorial letters (Flowerdew and Dudley-Evans 2002; Wahl-Jorgensen 2002, 2010) or sports commentary (Ferguson 1983; Reaser 2003). Not all of these studies are purely linguistic in nature, but other fields such as journalism studies and media research, despite not often using the term *register*, have contributed a lot to our understanding of these functional varieties. In contrast to the studies mentioned above which take a cross-variety approach to comparing several, unrelated registers, the study at hand aims at analysing variation within newspaper language and is therefore, despite including different regional varieties, confined to a specialised field in functional terms. This allows us to take into consideration the origins and traditions of mass media in the respective varieties to a much higher degree than studies comparing entirely different domains, and to draw on insights and results from other research areas for the interpretation and contextualisation of our linguistic findings.

In order to provide this background of mass media in general and newspapers in particular, the following section will take a look at the theoretical foundations in media studies and the aspects that have been of interest to scholars from various fields. It will then give an overview of linguistic perspectives on media language in more general terms before turning to the situations in the individual varieties, where basic data concerning media usage and newspaper circulations will be given. Finally, instead of being discussed at this point, research conducted into newspaper language of the individual regions will be described in the respective sections and thereby immediately placed into the broader context of the society and culture.

2.3 The Language of News

In his 1991 book, Bell assumes that "people in western countries probably hear more language from the media than they do directly from the lips of their fellow humans in conversation" (Bell 1991, 1). Independent of whether this estimation is true, media can certainly be regarded as being highly influential, on a linguistic as well as ideological level, and have therefore been analysed both theoretically and empirically from many different perspectives.

The phenomenon of mass media first attracted attention from the fields of philosophy and sociology, but gained importance also in political science and psychology in the first half of the 20th century especially with regard to its use as a means of spreading propaganda (see Bonfadelli 2004, 13). Although the actual influence of media consumption of any kind is almost impossible to determine, mass media impact has continued to be a dominant aspect in media research, and has become of interest to other fields of study as well when analyses were made to trace for instance the effectiveness of health information campaigns or environmental reports (e.g. Smith et al. 2002; Antilla 2005; Li et al. 2008). Since mass media not only influence their audiences, but also reflect them to a certain degree, they have been used for studies into societal issues such as gender representation, racism or creation of stereotypes (e.g. Meeks 2012; Aull and Brown 2013).

All of these studies provide very interesting answers to their respective research questions and contribute with their findings to the overall analysis of media, yet it would be beyond the scope of this work to go into detail about all the angles from which different media have been analysed. The following section will give a brief introduction to the area of media studies in general and trace the origins of scientific interest in the field and the theoretical developments in it; after this however, the focus will be narrowed down to linguistic approaches to news. In section 2.3.3, after providing this overview of linguistic interest, the attention will shift towards the newspaper landscapes in the concrete regions treated in this study, Australia, Hong

Kong, Kenya, the UK and the USA, and describe the most common research trends and methods applied to analysing the varieties of English in newspapers.

2.3.1 Origins and Boundaries of Media Studies

Media studies as a discipline in its own right is still fairly young, but long before its establishment and recognition, questions concerning media were asked and examined in other areas of research. As the concept of media and all its shapes were primarily considered a social phenomenon, it was especially the field of sociology which took an early interest in it. From this perspective, media was first of all a new factor to be considered in general theories of society; this led to media being examined not because of the subject itself, but because it had become a significant part of the overall picture. Leschke (2003, 163) goes so far as to say that media were used as a test field for theories and were not examined too closely lest the application should fail and thus falsify the whole theory.

This criticism is rather harsh, since in the context of a general theory, media can by definition only be one aspect. The advantage, also acknowledged by Leschke, of weaving media into an already existing theory is that it is immediately set into the broader context of society, which after all is the concern of sociology. A disadvantage can be seen in the fact that across the many works on the subject, media itself is often treated from the perspective of another issue which is the real focus of the analysis. One example for this tendency is Benjamin's (1991, originally written 1939) work, which discusses the transition observable in art that individual pieces of work such as paintings become reproducible due to technological advancements in photography and film. In this process, the reproductions are by no means entirely identical to the original; due to their wide distribution, they lose the context in which the original was produced and intended to appear, which in turn changes the perception of the respective item (ibid., 438-40). Different media thus constitute means not so much to change the piece of art itself, but the way it can be perceived and in this sense, they offer potential for entirely new forms of art (see also Mersch 2006, 64-70). A more critical view is taken by Horkheimer and Adorno (1947), who do not analyse media very deeply but take it into consideration because of the role it plays with regard to cultural aspects. Their evaluation of media, especially of mass media, is nowhere near as positive or optimistic as Benjamin's however, and they see it as the origin and primary means of what they call the culture industry, which degrades culture to the status of a product that is made purely for consumption.

This negative picture of media was shared by scholars especially from US-American contexts, while in Europe, there was a certain hope that the technology of mass media might bring with it new useful functions for society (see Mersch 2006, 59). This gave significance to the question of the impact media actually has on its audience and the public, an issue which had so far been considered theoretically by individual scholars particularly in England and France (e.g. Bryce 1888; Tarde [1898] 2010; Lowell 1914; see also Tsfati 2003), but had not been looked at empirically until in 1910, Weber suggested to establish what he termed the "sociology of press" in order to create a forum for analysing the then dominant mass medium of newspapers (Weber 1911, 42; see also Lazarsfeld and Boudon 1993, 293; Hepp 2004, 31; Schenk 2007, 3; Fest 2011, 13). Weber's intention was to separate this field from other areas so as to avoid overly strong influences from any perspective, however, the First and even more so the Second World War demonstrated the power of media as a means for propaganda and many of the studies that followed were therefore politically motivated, searching either for means to expose or to construct propaganda (e.g. Lazarsfeld, Berelson, and Gaudet 1944; see also Bonfadelli 2004, 13; Fest 2011, 13).

Apart from this, newspapers did not stay the only mass medium for long. The invention of the radio and the television in the twenties and thirties caused the focus of research to shift to these new technologies. In general, the history of media studies displays a trend that every new version of mass media dominated the research for a while and "has been hailed for its educational and cultural benefits, as well as feared for its disturbing influence" (McQuail 2005, 53; see also Bonfadelli 2004, 13; Fest 2011, 13). Additionally, the issue of mass media impact had raised interest not only in scientific circles, but became more and more a public discussion and was drawn on for debates on other topics, such as for instance media consumption of children or its triggering of violent behaviour. Empirical studies soon set out to examine the effects of media, however, many hypotheses and theories that were developed drew on the results of small statistical surveys or observations about mass media production (Fest 2011, 13). The accepted opinion was that newspapers, radio and television must have huge effects on their recipients as they were always present and constantly displayed opinions and perspectives to the world; "their omnipresence was taken as a proof for their omnipotence" (Fest 2011, 13; also Bonfadelli 2004, 28). This understanding of mass media was underlined by a number of smaller studies, like for instance Welles' 1938 radio play "Invasion from Mars". The show, which had been intended as a practical Halloween joke, was broadcast to interrupt the running programme on CBS to announce that Martians had invaded New Jersey, and as it was produced like a live-documentation, featuring soundbites from journalists, police officers and civilians, it was taken as genuine news and caused a mass panic that spread before the play was over (see also Fest 2011, 13). This accidental experiment was regarded by many scholars to confirm the basic stimulus-response-model which assumes that a carefully constructed stimulus spread via mass media reaches every recipient in the same way, thus triggering the same reaction and leading to mass movements (ibid.; Schenk 2007, 24).

This model is of course very limited and simplified, but in the following years, the technological advancements that changed the landscape of mass media also improved the means for empirical studies. Beginning in 1940, Lazarsfeld, Berelson and Gaudet conducted a series of experiments to analyse the influence mass media consumption had on the political opinions of its recipients, and found that the effects were surprisingly small. Instead of being influenced by what they heard or saw on media channels, the researchers came to the conclusion that people were much more affected by direct conversations with peers and felt to receive more, and more frequent, information from this context as well (Lazarsfeld, Berelson and Gaudet 1944, 150; Lazarsfeld and Menzel 1963, 96; cf. also Schenk 2007, 350-51). This notion that so clearly contradicted the assumption of omnipotence of mass media was strengthened further by studies conducted into potential effects of violence on television and in films, which aimed at explaining violent behaviour of recipients, especially teenagers and children (Bonfadelli 2004, 22-23; Fest 2011, 14). Early tests showed a less significant relation between these two factors than expected and instead led to the conclusion that media impact is more complex than a simple chain of cause and effect and depends on context, attitude and subjective reception (e.g. Zillmann 1979).

With assumptions about mass media thus ranging from one extreme to the other and the actual subject, namely effects, being so hard to measure, Klapper (1960) took up the results from Lazarsfeld's study and suggested to treat mass media not as a highly effective and manipulative institution but as one influence out of many people encounter in their surroundings. In accordance with Lazarsfeld, Berelson and Gaudet, Klapper comes to the conclusion that recipients of mass media go through processes of “selective exposure”, choosing their sources of information based on their existing opinions, “selective perception”, paying particular attention to information that confirms their beliefs, and “selective retention”, remembering mainly what fits into already acknowledged patterns of facts (Klapper 1957, 459; 1960, 19-25). This concept is rooted in the framework of social psychology and offers an expansion of the model of cognitive dissonance presented by Festinger (1957), which describes the negative emotions triggered if a person encounters subjectively incompatible ideas. Such a discrepancy will lead to discomfort, and the respective recipient can attempt to change this either by adding consonant elements, by decreasing the relevance of dissonant aspects or by reinterpreting the elements and re-categorising or changing what appears dissonant (see also Fischer and Wiswede 1997, 227-43; Fest 2011, 14). Klapper’s expansion “adds to this a sort of pre-stage in which a person seeks to avoid the uncomfortable sensation of dissonance from the beginning” by selecting consonant content from the start (Fest 2011, 14).

Klapper’s model reached its limits when it was confronted with the increasingly complex landscape of TV which characterised mass media development throughout the seventies, eighties and nineties. The idea of selection as described by him was counteracted by the observation that a recipient is not entirely free, but can only choose content from what has previously been filtered by media producers and journalists, and that the huge amount of news and information offered by the various mass media channels, especially on television, precludes the possibility for a recipient to collect all information and to protect themselves from dissonant opinions (Noelle-Neumann 1995, 546; Schenk 2007, 764; Fest 2011, 14). This increasing complexity made potential effects of media consumption even harder to determine, especially in combination with the fact that despite technological improvements, the respective highly psychological processes were, and still are, very hard to grasp (Fest 2011, 14).

One reason for this, offered by Bonfadelli (2004, 12; see also Fest 2011, 14), is the one-sided nature of communication via mass media: the possible audiences of television, newspapers and radio are huge as well as diverse with regard to multiple factors, which makes the observable effects fleeting and elusive and therefore hard to determine. This view is shared by Luhmann, who stresses the importance of this one-sided nature of communication which does not permit any interaction between the producer or the journalist on the one and the audience on the other hand (Luhmann 2000, 2). The relationship that exists between the discourse participants is not equal, and the only possible way for recipients to interact in connection to received mass media input is a discourse amongst each other; this, however, does not have any immediate impact on the journalist or producer: "Of course, oral communication is still possible as a reaction to things which are printed or broadcast. *But the success of scheduled communication no longer depends upon it*" (ibid., 16; emphasis in original; see also Fest 2011, 14). This is true for the 'older' forms of mass media, namely TV, radio and newspapers, but has changed since.

The complexity of the field of media has increased steadily, and new forms, most of all of course the internet and the different forms of communication it implies, have arisen and taken the concept of media to a whole new level. In this context, the relationship between producer and recipient has also undergone alteration; the production of news items published by TV or radio stations and newspapers still lies entirely with the journalist, but once the content has been published, recipients can react to it within seconds via various channels, be it the company's or even the individual journalist's website, their respective Facebook-pages or Twitter-feeds. Furthermore, the communication between the recipients can have unpredictable effects due to the fast and easy channels offered online, resulting in phenomena like so-called shitstorms or news going 'viral' and spreading across various platforms and forums far beyond their original source or intended scope. The potential of the online world also challenged the traditional understanding of the very roles of producer and recipient, as everybody can publish news and information online with just a few clicks, thus switching between the roles constantly.

These new forms and their production processes are often discussed under the term of "convergence media" (e.g. Lanson and Fought 1999; Lewis 2003; Franklin et al. 2005; contributions in Quinn and Filak 2005; Kolodzy 2006), and the developments have brought about massive challenges not only for reception research, but for the whole field of media studies in general. Since the current study focuses exclusively on newspaper articles written by journalists however, and the individual papers display very different standards as to the technological implementations of their websites and the functions they offer, these aspects will not be an explicit part of the study. Furthermore, a brief description would hardly do these emerging research spaces justice; this should not be taken to mean that they do not influence newspaper journalism; on the contrary, being online has allowed newspapers to be much more flexible with regard to article length and publication time (Lewis 2003, 97). Traditionally, daily newspapers were doomed to cover news from the day before as they were printed during the night to be out in the morning. The printed versions of newspapers still have this restriction, but the online presence that almost all papers by now have established means that in addition to what is put in print, numerous articles can be published at any time and with a negligible delay between production and publication. This affects most of all the component of recency of news, as newspapers can now react to events as fast as for instance radio programmes. This can be expected to change the temporal scope still counting as "recent" and thereby altering the story's news worthiness in this respect (cf. Bednarek and Caple 2014; Potts, Bednarek and Caple 2015). This change in infrastructure triggered by newspapers going online should therefore be kept in mind when discussing newspaper language.

Another feature that the launch of online versions of newspapers has brought with it is that of comment functions. In principle, these offer readers the possibility to comment on articles, but there are different ways these tools are handled. While some are actively moderated by staff members, others are left open and at best simply delete comments with critical content. Also, although these comments theoretically have the potential to create a direct communication between recipient and producer, the journalists do not often get involved in the discussions in practice and interactive elements are still limited, though increasing (Schultz 1999; Singer 2006). Several studies have been conducted into the language and function of these comments, looking for instance at the effects of anonymous communication (e.g. Santana 2014), the role comments play for regional and local newspapers (e.g. Canter 2013) or the acceptance of this user-generated content among journalists and other readers (e.g. Hermida and Thurman 2008). An effective examination of these comments necessarily includes a coupling of the comments with the respective article they refer to as well as a look at intertextuality between the comments to determine the argumentative structure of individual contributors and development of thematic foci. Since this would go well beyond the scope of this work and not all newspapers included in the corpus of this study feature comment sections on their websites or restrict them to certain articles only, this will not be taken into consideration in the analysis of the data.

One obvious effect the rapid development in the field of media has on every research area interested in it is one of terminology. The problem of what *media* effectively means has been an ever present one in the field, and is not reduced by new forms coming into existence or changing so quickly. The concepts of media as presented by Luhmann or Bonfadelli are certainly too narrow for today's world, yet a definition that is too broad turns it into a term that is impossible to work with. An example for the latter is McLuhan's definition, which describes a medium as an "extension of ourselves" (McLuhan [1964] 2001, 7). He goes on to say that "the 'content' of any medium is always another medium. The content of writing is speech, just as the written word is the content of print, and print is the content of the telegraph" (ibid., 8). This gives an idea of the difference between his definition and that used by others – while many scholars use *media* as an uncountable noun, McLuhan's concept can be traced back directly to the singular *medium*. His notion is still wide and includes everything "from the wheel to the computer" (Leschke 2003, 11; my translation), but it shows by contrast that more commonly media is regarded as something abstract, a concept rather than a list of clearly defined individual instances. The level of abstraction has certainly been raised by the recently developed and constantly evolving internet options, leading to an even harder definition of the term *media*.

As was said above, this dimension is not the focus of this work; neither does it claim to cover the terminological discussion, let alone offer a solution for it. The main aspect being of linguistic nature and narrowed down so clearly to newspaper language, it would be unreasonable to add to the terminological confusion in this respect, which is why we will employ the terms *media* and *mass media* synonymously and in as simple a sense as is still workable for the purpose of this work, namely referring to "the main means of mass communication, such as television, radio, and newspapers, considered collectively" (OED 2016), including both traditional broadcasts as well as online publishing of these media. Sporadic use of the term *medium* is not to indicate a lesser degree of abstraction of the concept, but to be seen simply as a reference to one particular mass media type.

2.3.2 Linguistic Perspectives on Media Language

With general media theories aiming at embedding media as a social phenomenon into the broader context of society, the analysis of the language used to transport medial content has become a rather specific subdomain of media studies. Meanwhile, it has received much more attention from the field of linguistics as media language has been found to be highly influential on its recipients (Fowler 1991, 7; Conboy 2007, 4, 175; Bednarek and Caple 2012b, 6) and therefore also to reflect linguistic characteristics of target groups and audiences (Bell 1984a, 74). With the background of media theory and reception research as outlined above in mind, this section will briefly describe the research that has so far been conducted regarding the language of media and the methods that have been applied, putting an emphasis on newspaper language.

One notable distinction among works that deal with this subject is that some are written for academic purposes while others are of a more practical disposition and aim at the producers of media content. The latter are usually very prescriptive in their approach, and are often intended as guidebooks or manuals for aspiring journalists (see also Fest 2015, 55). Nevertheless, their content is valuable to a linguistic analysis of media language; by explaining how for instance news items are to be written for different media, they become to a certain degree the source of linguistic characteristics of these stories. Although individual broadcast stations also implement their own rules and often have style guides for their staff, it is therefore necessary to be aware of the most common stylistic devices presented in independent prescriptive works and textbooks.

The number of volumes that have been published to this end is nearly infinite, and not all of them will be recounted here. Most of the works in question specialise on one particular mass medium, as the requirements are very different. For TV work, it is essential for a journalist to keep images and text in a balance, and the two condition each other (for a closer analysis of different models for text-image relations, see Burger 2005, 400-24). Either of them can provide the pivotal point of an item; ideally, both images and storyline are adequate for a news item, but when the images available for a story are impressive, but the story itself offers little content, or if the story is relevant but no good images can be found, it might still feature in a TV show (Hagen 1994, 209; Ekström 2000, 468-77). The visual aspect in itself can thus be a criterion for news worthiness which is non-existent in radio or newspaper work. What TV does share with radio broadcast is the non-permanence of its content; a hearer or viewer only gets the content once, and the speed for the succession of items is controlled by the producers, not the recipients (Katzenberger 1999, 91). Although both media nowadays often feature online streams, podcasts or, in the case of TV, video on demand, which allow the recipient to go back to individual items and replay them, both still mainly have to cater for a target group which listens or watches live: "The listener or viewer *could* rewind the programme to listen again, but that requires effort. […] The onus on making sense of the news always lies with the newswriter and newsreader, never with the audience" (Boyd, Stewart and Alexander 2008, 71; emphasis in original). Especially TV has been found to be a typical medium consumed on the side (Kolodzy 2006, 13), and both media can be assumed to have high fluctuation rates in their audiences.

The language has to adapt to this, and the guidelines offer very clear suggestions. Sentences should be shorter than in written language, with less complex syntactical structures. Any kind of ambiguity is to be avoided, as are technical terms or rare expressions (Boyd, Stewart, and Alexander 2008, 71-77). In general, writers are advised never to use a long word when a shorter one is available and to leave out redundant words (Boyd, Stewart, and Alexander 2008, 93). Especially in news, clichés and overly exaggerated statements are considered critical as they make an item appear less serious. Furthermore, features like figures, acronyms and homonyms have to be treated with care as they require more time to be processed by the audience and can therefore lead to confusion (Kolodzy 2006, 135; Boyd, Stewart, and Alexander 2008, 94-104).

The situation is slightly different for newspapers. First of all, an average newspaper consists of actual news to a greater degree than most TV or radio programmes; the former contain, depending on the channel, talk shows, live broadcasts, sitcoms, movies and magazine shows, and the latter contain varying amounts of music and verbal contributions in the form of interviews, expert talks or studio guests, to name but a few. News is one fixed element out of many, while in newspapers, in contrast to magazines as another print medium, it is the main part. Furthermore, reading a paper is not something that can be done alongside another activity, as it demands more attention. It does not have to consider any time restrictions concerning broadcast schedules either, however, at least the print versions are limited in space (Fest 2015, 54-55).

These external characteristics of course influence the language in use. The syntax used in newspaper items can be more complicated than on radio or TV, because readers can read a sentence again or jump back and forth in the article as they please. However, a reader will give up when a text is too tedious to read and understand, which puts limits to the complexity of syntactical constructions (Lüger 1995, 23; Harcup 2005, 107-8; Kolodzy 2006, 157; Fest 2015, 55). Despite the increased control on the side of the recipient, there are therefore rules to follow as well. On the levels of lexis and syntax, many similarities can be found to what is recommended for television and radio writing, namely not to include abbreviations that might be unknown, to avoid too much technical terminology and to include as few clichés as possible. In newspapers, too, active is to be preferred over passive voice, and the language is to be simple and to the point. However, in contrast to television or radio broadcast, newspaper articles often work with puns as it is less problematic that the audience might need more processing time (Pape and Featherstone 2005, 49-56; Keeble 2006, 96-97).

Also, ambiguity can be avoided more easily in newspaper articles, especially with regard to quotations; although quotations are considered helpful in broadcast, it has to be taken into account that the audience cannot see punctuation. Unless the quote is given by the original person directly, as a sound bite or a short video sequence, including it read by the moderator or journalist themselves can lead to unclear statements. Boyd, Stewart, and Alexander (2008, 73) exemplify this with the following sentence:

(1) 'Ethiopia said the Eritrean leader had started the conflict.'
(2) 'Ethiopia,' said the Eritrean leader, 'had started the conflict.'

While in print the sentences are clearly distinguishable and perfectly unambiguous due to punctuation, they could not be told apart easily without this visual aid. Newspaper articles therefore have more freedom concerning the inclusion of quotations and can be expected to feature more than broadcast news.

On a structural level, the assumption that a newspaper reader wants to get informed in as little time as possible is the dominant influence and implies that the headline and the first paragraph of an article should bring across the most vital bits of information on the story. The following parts of the article then give more detail and provide the background, resulting in the so-called "inverted triangle" of information density (Randall 2000, 147; Harcup 2005, 108-13). As a rule, a reader should be able to stop reading at any point and still get the main aspects of the story, even the headline should function as a news bulletin when standing on its own (Burger 2005, 114). Throughout the article, the questions *Who?*, *What?*, *When?*, *Why?*, *Where?* and *How?* should be answered if possible (Keeble 2006, 111).

These rules are not universal and their applicability may vary depending on the story, the news domain or the paper, however, they reflect basic assumptions about the function of newspapers and the demands of readers. Finding these suggestions realised in the language should therefore be traced back to a considerable degree to their origin in prescriptive guidebooks for journalists and the purpose of newspapers. In contrast, deviances from these norms, especially if they turn out to be characteristic for any particular grouping (for instance a particular paper, country or topic), can offer insights into varying standards or purposes within the field of newspaper language.

From a scientific point of view, across different mass media newspapers have attracted the attention of particularly many scholars. One reason for this certainly is that in contrast to TV and radio broadcasts, newspaper texts are comparatively easy to collect and to process, as no transcription is necessary. Thus in many studies they are taken as one of the most popular and widespread examples of public written language and are included rather as a representative for that than with the aim to analyse newspaper language in particular (see e.g. Biber et al. 1999; Biber and Conrad 2009). This is not to say that these studies do not contribute to the understanding of this language variety, quite on the contrary, they often place newspaper language in comparison to other functional varieties and therefore filter out characteristic linguistic features very effectively. The descriptive results are valuable not least as comparative values for any other study, but also because they give a first impression with regard to where in the almost endless network of functional language varieties newspaper language can be located. However, when discussing the findings they do not usually take into account background knowledge from non-linguistic areas, often leaving conclusions to be drawn as to the actual meaning of the results for the understanding of media.

Other studies look at newspaper language in an isolated state, and follow research aims that are either directed at the functional variety directly or at sociological issues that are reflected in it. As was said above, the language used in mass media publications is very influential and far-reaching, and newspapers offer themselves to these studies not just because of the methodological advantages described above but also because in newspapers, the language plays a more crucial role than for instance in TV, where the visual input is just as relevant. Newspapers do feature other stylistic devices as well of course – articles are often accompanied by pictures, text passages, especially headlines, are formatted differently and online versions of papers offer the possibility to embed video material as well. Individual aspects as well as the interplay of these multimodal devices have been the focus of several studies (e.g. Schneider

2000; Bateman, Delin and Henschel 2007; Knox 2009a, 2009b; Caple 2010), but especially for quantitative work the implementation of all these multimodal layers in one analysis is complicated. As the newspapers in the corpus for this study use very different means and technological resources, the aspect of multimodality will not be part of the analysis in order to avoid incomplete results and, based on these, flawed conclusions.

Those studies that look at the language in use display a vast variety of research foci. Some can be clearly assigned to a particular approach and the therein implied questions. Especially CDA, with its focus on power relations, offers itself for ideological research questions and of course draws on media content for many analyses due to its wide reach and influence. It is used for instance in KhosraviNik's (2009) study which analyses the issue of refugees and its representation in newspapers, in Stamou's (2001) analysis of the portrayal of non-protesters during protests and demonstrations in Greece, or by Teo (2000), who looks at reporting of racism in Australian newspapers. Broader applications of CDA to news language are offered by van Dijk (1983, 1984), Fairclough (1995) and Richardson (2009), of which especially the latter frequently draws on journalistic concepts and thus offers a diverse, yet mainly introductory, perspective on the subject. Furthermore, several studies have applied the more general framework of discourse analysis to analyse the linguistic representation of news values, or newsworthiness (Bednarek and Caple 2012a, 2014; Caple and Bednarek 2015; Potts, Bednarek, and Caple 2015), using both quantitative and qualitative approaches.

Many of these studies have very clear-cut research goals that go beyond the purely journalistic level. Mass media are not produced randomly but reflect their producers as well as recipients, and can therefore be seen as a "snapshots of our life and our culture" (Reah 2003, 1). It is therefore not surprising that the analysis of news language has found significant resonance also in the field of sociolinguistics. The above-mentioned studies drawing on CDA of course have sociolinguistic dimensions and implications, but power – and in that context often also politics – are just one angle to take. Many sociolinguistic studies of news discourse take a particular interest in the connection between and mutual influence of producers and audience and relate the assumed target group of for instance a newspaper to the style used in its articles (e.g. Bell 1991; Jucker 1992). This concept, which Bell (1984b, 145; 1991, 121) calls "audience design", is closely connected to approaches from media studies and can be seen as the media's side of Klapper's idea of the recipients' selectiveness. Other sociolinguistic studies take this connection between medium and audience into account and analyse it for particular media or domains. Two such examples are the works of Conboy (2006), who looks at the construction and representation of target groups in British tabloid newspapers, and Crawford (2009), who analyses the inclusion of the audience in the opinion sections of Scottish quality press. These studies usually draw on journalistic knowledge to a great degree and take into account the prescriptive guidelines described above to include an image of the production processes of news and its influence on the language of the final output.

Another approach to determine characteristics of newspaper language is the systemic functional one. In this context, especially the framework of Appraisal has found a huge field for application. Vestergaard (1999, 89) observes that "the events reported in newspapers should of course be both new and true, but they should also appear to be true". By default, news is often taken to be factual and objective by the audience and clearly evaluative instances of language, most of all any that explicitly give the author's own opinion, are rare. Nevertheless, news is of course exposed to numerous subjective influences; the selection, the place and length the items are allocated in the paper as well as editing by various people involved all contribute to the final product just as the topic itself does (Gilles 1997, 69; Conboy 2007, 9). Many studies (e.g. Bednarek 2006, 2008; White 2000, 2004, 2006) have therefore been conducted to analyse instances of evaluative language in news. Other works take a register- or genre-based[14] approach to the language of news and have tried to define characteristics either for news language in general or for individual news domains, taking appraisal, but also other linguistic features into account.

In this sense, van Dijk (1985), White (1997) and Mahlberg and O'Donnell (2008) look at typical text structures of hard news in English, and Thompson, White, and Kitley (2008) analyse structural elements in hard news in English, French and Japanese. Reinemann et al. (2012) compare hard news to soft news and examine the respective styles in order to reach a more clear-cut distinction between the two. With regard to topics from economy or politics, studies usually follow narrower research aims than to describe the language in use, often connected to issues of power and manipulation

[14] Again, the terminology varies greatly between the studies, but whichever of the two terms is in use refers to one specific domain of news, usually in one particular medium.

(e.g. Carvalho 2005; Collins et al. 2006; Hopkins 2011). Features or lifestyle articles have been analysed surprisingly rarely. The most notable empirical studies are analyses by Welch, Fenwick, and Roberts (1997, 1998) dealing with feature articles on crime-related topics, and a comparison of lifestyle articles in different varieties of English (Fest 2016); otherwise, they are usually only mentioned in passing in manuals (Friedlander 1982; Granato 2002; Ricketson 2004) or prescriptive journalism textbooks as those mentioned above. In contrast to this, the field of sports news has received particular attention, partly because of the social implications of sports, which have often been traced in the language used to represent it in the media (see e.g. Ghadessy 1988; Beard 1998; Eastman and Billings 2000; contributions in Lavric et al. 2008; Dreyfus and Jones 2010; Fest, forthcoming), and partly because it is so clearly distinguishable. In all mass media, be it online news, television, radio or newspapers, the sports section is usually clearly marked and often separated from the rest, and news items are easily recognisable as belonging to this category. Other thematic domains are harder to define as their topics overlap more and they are often not marked by the producers. As a result, the labels "hard news" and "soft news" are often applied but are hardly ever used consistently. The study at hand distinguishes the domains of economy, politics, lifestyle, sports and hard news, and more clear-cut definitions will be given in section 4.2.2.

In many studies, the question of differences not only between news domains but also between countries and cultures has been raised. However, the two perspectives are hardly ever combined, as almost all work focuses either on a comparison between news domains, not taking into account differences between languages or regional varieties of one language, or on a comparison between languages on the basis of news language in general or within the scope of only one domain. Yet the close connection between news and the societies for which they are provided calls for such a cross-dimensional perspective: in order to truly get a grasp on the culture beyond, it is not just newspaper language in general but the degree of differentiation between the individual topics and the therein reflected relevance attributed to them that is of interest. Furthermore, such an analysis of newspaper language for a specific regional variety is best conducted in comparison to others, so that both regional and functional differences are traced. This is important in order to assign linguistic characteristics to the right source; if a variety displays few differences between two particular thematic domains, this does not necessarily entail typicality for the regional variety, but the similarity can be characteristic of the two domains in general, thereby rendering the respective regional variety as not outstanding at all. The two-way comparison across regional as well as functional varieties conducted in this study thus has the side-effect of functioning as a test for assumptions in either direction.

In order to understand and interpret results from a news analysis correctly, it should be taken into account that different societies have different media structures and landscapes. The regions discussed in this study vary greatly in size and population numbers, and consequently also in the amount of newspapers that are published and their respective scopes. Furthermore, different traditions can be traced with regard to the cooperation between papers and the constellation of media companies. To ensure sufficient background information for the following analysis, the next section will therefore provide an overview of the newspaper situations in the five varieties included in the corpus.

2.3.3 Approaches to Different Newspaper Landscapes

Newspapers in general are the oldest mass medium. They developed long before the radio and television and are present in one form or another in almost every society. They can take different forms of course; there are daily newspapers, weekly and monthly newspapers, special Sunday editions and evening papers. Some are published in the so-called broadsheet format while others come in Berliner format or tabloid shape, to name just a few. With the rise of the internet, online editions became important publishing formats and most newspapers present their content online either for free or via subscriptions.

Despite these general trends, every society displays a unique newspaper landscape that is influenced by a number of factors. Mass media did not develop simultaneously or identically all over the world, but in connection to the demographic and political development of the society itself. In some areas, especially those with a colonial background, it was heavily controlled for a long time; in other regions it was eyed with suspicion due to the influence mass media consumption was believed to have on people while it was at the same time acknowledged to be the most effective way to spread news and keep the public informed (see section 2.3.1). As a result, different degrees of censorship evolved, from merely advisory forms like age recommendations all the way to extreme, usually politically motivated, control of content, which is still reflected in the World Press Freedom Index today (cf. Reporters without Borders 2015; see Figure 2.5).

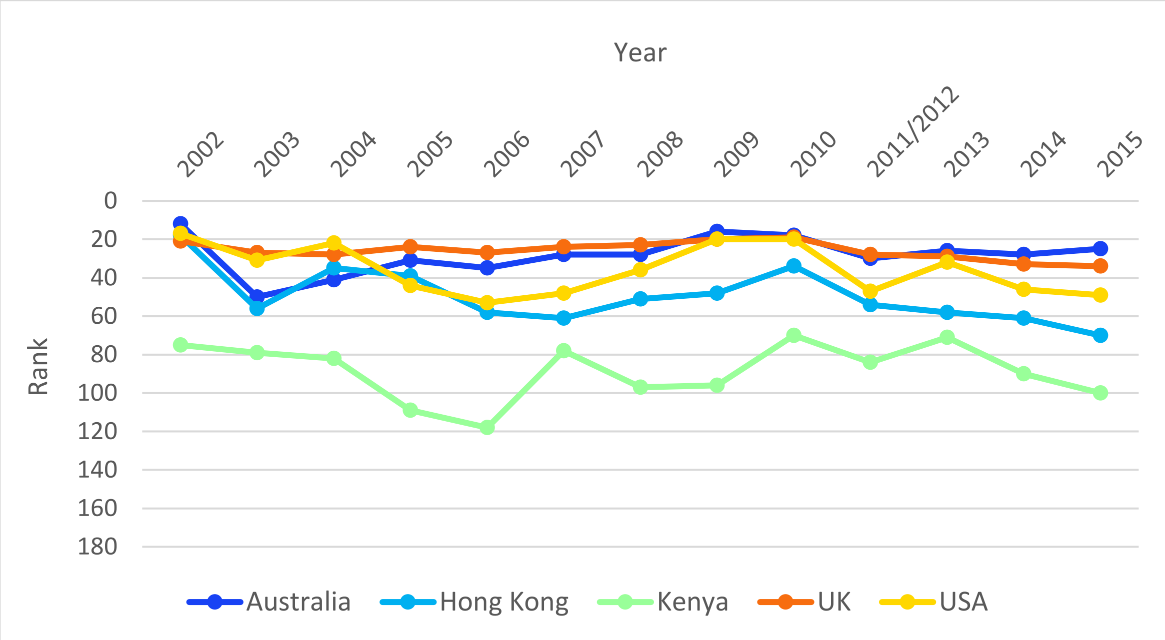


Figure 2.5: World Press Freedom Index - Development

Apart from ideological influences on media, especially television and online media require a technological infrastructure that is not necessarily given everywhere. In these regions especially, newspapers can be expected to play a more important part than elsewhere, while at the same time their online versions, if at all available, might be used comparatively little. It is the aim of this section to give an overview of the situations in Australia, Hong Kong, Kenya, the UK and the USA regarding media availability, accessibility and usage,[15] keeping a focus on newspapers as the medium in question. Furthermore, the subsections will give brief descriptions of the research that has been done on newspaper language in the respective regions.

[15] Although the subsequent sections will try to describe newspaper circulations as accurately as possible, reliable figures are not available for all regions and papers and are often estimates rather than approved data. Where figures are available, they do not necessarily refer to the same month or year; furthermore, in some cases, Sunday or weekend editions of newspapers are measured separately while for other papers, such detailed data is not available. Similar problems occur with online editions: for papers which are not freely available, figures for paid subscriptions are sometimes provided, while for free papers, the numbers of online users cannot be determined except by amounts of clicks and browsers. The tables in the following sections will – as far as can be told – refer to the weekday editions of newspapers and will not include potential figures for online subscriptions so as to keep the lists as comparable as possible.

2.3.3.1 Australia

Of the five countries included in this study, Australia is by far the least densely populated with an average of 3 people per km^2. It covers an area of 7.69 million km^2 and in 2014 had a population of almost 24 million people, 14.9 million of which lived in and around the five biggest cities of Sydney, Melbourne, Brisbane, Perth and Adelaide (Australian Bureau of Statistics 2015, 6, 18; Australian Government - Department of Foreign Affairs and Trade 2015). The literacy rate in the country is considered >99% (Ethnologue 2016), and in 2011 more than 99% of all households had access to television and radio broadcast, with the majority of households (73%) having two or more TV sets (Australian Communications and Media Authority 2012, 6). In 2014 an estimated 89.6% of Australians had an internet connection (CIA 2016).

In terms of technological infrastructure, Australia is therefore well equipped for all types of mass media. Yet the particular demographic structure of the country, especially the huge area it covers in relation to the widespread, yet in spots highly concentrated population, has a huge influence on the media landscape. Australia consists of two territories, the Northern Territory (NT) and the Australian Capital Territory (ACT), and six states, New South Wales (NSW), Victoria (VIC), Queensland (QLD), Western Australia (WA), South Australia (SA) and Tasmania (TAS). These regions vary greatly in size, but all except the Capital Territory are still significantly bigger than for instance Belgium or the Netherlands. It is therefore not surprising that with regard to newspapers, local papers dominate in numbers as they can more effectively cover the region that is of immediate relevance for the audience than a publication which aims at covering all of Australia, and all eight regions have numerous papers of different scales, ranging from community ones to state-wide issues. The two biggest national daily newspapers are *The Australian Financial Review* and *The Australian*, however, many of the regional ones are also published nationally which renders the boundary between national and regional newspapers rather blurry.

Independent of the type of paper, newspapers still hold a very high relevance in Australia. In a 2014 survey 91% of Australians stated that they read newspapers regularly, with 29% of the readers drawing only on print versions, 16% solely on digital editions and 55% on both (The Newspaper Works 2015, 5). Metropolitan and nationally published papers have the largest audiences, and especially sections focusing on specific topics, such as arts, property, travel, business and sports, are very popular (ibid., 7). The circulation numbers in Australia are below those in many, much smaller countries like for instance the UK, but since these figures are absolute values this can again be traced back to Australia's comparatively low population and numerous local newspapers. They also reflect the tendency towards regional rather than national papers: the papers with the highest circulation numbers are the *Herald Sun* (VIC), *The Daily Telegraph* (NSW), *The Courier Mail* (QLD), *The West Australian* (WA), *The Advertiser* (SA) and *The Sydney Morning Herald* (NSW), all of which focus on their respective area but are published nationwide and therefore cover national and international news as well. The national papers on the other hand reach lower circulation numbers.

Despite the wide range in newspapers and their scope, the picture of ownership is much less diverse. Almost all large papers in Australia are part of either News Limited or Fairfax Media, which is reflected in the articles they publish as the papers share content and often rely on the same authors as well. Of the twelve national and regional papers, seven belong to News Limited and four to Fairfax Media. *The West Australian* alone stands out as being owned by WA Newspapers (Audit Bureau of Circulations 2013, 1-2; see Table 2.1).

Table 2.1: Average circulation of English-language daily newspapers in Australia (national and metropolitan), Mon-Fri editions

Newspaper	Origin/Scope	Circ.	Owner/Publisher
Herald Sun*	regional (VIC)	399,638	News Ltd
The Daily Telegraph	regional (NSW)	289,839	News Ltd
The Courier Mail*	regional (QLD)	172,816	News Ltd
The West Australian*	regional (WA)	172,188	WA Newspapers
The Advertiser	regional (SA)	148,430	News Ltd
The Sydney Morning Herald*	regional (NSW)	136,623	Fairfax Media
The Age	regional (VIC)	133,981	Fairfax Media
The Australian	national	116,854	News Ltd
The Australian Financial Review	national	64,270	Fairfax Media
The Mercury	regional (TAS)	35,821	News Ltd
The Canberra Times	regional (ACT)	25,228	Fairfax Media
Northern Territory News	regional (NT)	16,757	News Ltd

(Audit Bureau of Circulations 2013); *newspapers included in this study

In order to ensure as much variation as possible in the corpus for this study and to avoid too much text produced by the same authors, we decided against simply including the four largest papers, although all of them would have been freely accessible online. Instead, *The Daily Telegraph* was left out and replaced by *The Sydney Morning*

Herald, the next non-News Limited-paper in line. Although the *Telegraph* has a slightly higher circulation than *The Courier Mail* and would therefore have been preferable to include on these grounds, the latter represents a different region, Queensland, while the *Telegraph* is published in Sydney, like *The Sydney Morning Herald*. As many of the articles published in the two papers are shared, this regional aspect outweighed the differences in circulation.

Although the relations between papers in terms of ownership should thus be kept in mind in the analysis due to the issue of shared content, it should also be noted that despite the rather uniform corporate landscape, Australia was registered with a higher degree of press freedom than the other countries included in this study in the current World Press Freedom Index, and is also the only one of the five that improved in this aspect in comparison to last year's ranking by jumping from rank 28 to 25 (Reporters without Borders 2015). Consequently, aspects like censorship and political influence on media, which are dominant throughout the research into for instance Hong Kong news (see section 2.3.3.2), are hardly considered at all with regard to Australia. Instead, a number of studies deal with news and media campaigns on health issues such as smoking (e.g. Chapman and Dominello 2001; Durrant et al. 2003; Clegg-Smith et al. 2005), skin cancer (e.g. Smith et al. 2002), HIV (Lupton 1993) or aspects of malnutrition (Li et al. 2008). In these analyses, the language of the respective news items is not the focus; rather, the predominant question is whether news coverage of these issues has any influence on the behaviour of people and helps to raise awareness. A similar focus on media impact is taken by studies analysing news reports on suicides and their potential connection to an increase in suicides in the population, approaching the question for instance from a medical (Pirkis et al. 2006) or a psychological (Hassan 1995) perspective. A more sociological point of view on media impact is furthermore taken by Dowding and Lewis (2012), who analyse our conception of corruption in politics in relation to the amount of articles published on the topic.

These research ideas, which build on a societal foundation rather than an interest in news language as such, can be found regarding other topics as well. The representation of gender, especially women, has been analysed by Stirling (1987) and Ryan (2005) and, with a particular focus on sports news, also by Kane and Lenskyi (1998) and Wensing and Bruce (2003). In these studies, a qualitative approach is applied, while papers analysing environmental issues and science in the news appear to prefer a quantitative approach. Hurlimann and Dolnicar (2012) evaluate the amount of news coverage on issues relating to water, such as fresh water reserves and water pollution, and Bubela and Caulfield (2004) analyse whether articles concerning genetic research can be said to represent the topic adequately or tend to create a hype around it. Both analyses originate from non-linguistic backgrounds and concentrate on their respective fields, environmental research and medicine, rather than language. In contrast to this, Bednarek and Caple (2010) examine the representation of environmental issues on a linguistic level, particularly taking into account the multimodal layout of newspapers.

One other regularly researched societal topic, which brings with it also ideological dimensions, is that of the representation of different nationalities or ethnic groups within or outside of Australia. A very straightforward example is Conduit's (1991) study which compares the public opinion on Indonesian people and culture with the picture painted in news. Since Australia has always been a popular destination for immigrants however, it is no surprise that especially the question of how asylum seekers are depicted in the news has been researched multiple times: Pickering (2001) and Macken-Horarik (2003) both analyse it qualitatively, the former drawing on content from *The Courier Mail* and *The Sydney Moring Herald*, the latter on *The Daily Telegraph*. Klocker and Dunn (2003) take their data from *The Advertiser* and *The Sunday Mail* and, taking a quantitative approach, compare their news coverage to official press releases from the government to trace differences and indicators of mutual influence. Despite the fact that asylum seekers and immigrants are often treated as generic and anonymous groups in the press, other studies have looked in more detail at the representation of smaller groups, such as Sudanese Australians (Nolan et al. 2011) or Muslims (Manning 2003). Especially the second study represents a trend in research that started after the attacks on the World Trade Centre in 2001, although it is often focused on news coverage in the USA rather than in other countries (see section 2.3.3.5). An ethnic group more particularly represented in research into Australian news are Aborigines, and their role in the Australian public is described both by analyses into Aboriginal newspapers (e.g. Avison and Meadows 2000) and into their depiction in (Meadows 2000) and linguistic influence on (Leitner 2000) English-language news.

Apart from studies which take ideological or societal foci, Australian media itself has been analysed from various angles. About some parts of the country, there are historical descriptions about the development of media, the most notable ones concerning Sydney (Isaacs and Kirkpatrick 2003) and North Queensland (Manion 1981). Furthermore, the concept of local and community papers has attracted a lot of research interest in Australia, as their importance, as was described above, is comparatively big in such a large and yet very loosely populated country (e.g. Ramson 1989; Kirkpatrick 2001; Bowd 2003; Rojas-Lizana 2011). Different types of newspapers are also considered by Flew (2009), who looks at the degree of interactivity in online papers, and Henningham (1996), who characterises the different content priorities set by popular and quality papers.

Another distinction that can be observed in the literature is that of specific news domains. In this context, especially hard news receives a lot of attention and is analysed based on for instance its structure (e.g. Iedema 1997), its multimodality (e.g. Knox 2009b), its reader-author-relationship (e.g. Scott 2011) and the news values it dominantly constructs (e.g. McKay 1983). Other domains of interest are comment sections (e.g. Wang 2008), economic news (e.g. Zhu and McKenna 2012) and obituaries (Starck 2004). While the latter compares Australia to the USA and the UK, most other studies in this context of individual domains are restricted to Australian news only, and usually make use of both qualitative and quantitative elements in their analyses. With regard to studies looking at linguistic features on the lexico-grammatical level, a qualitative approach is more common; Turner (1972) for instance analyses grammatical constructions in headlines, drawing mainly on *The Advertiser* and *The Australian*, and Peters et al. (1988) look at the use of adverbials, conjunctions and other lexical items in various newspaper articles from different sources. Only Estling (2000) examines grammatical constructions using a corpus-based approach, and comparing Australia – represented by *The Sydney Morning Herald* – to the UK (*The Independent*) and the USA (*The New York Times*). Especially Estling (2000) and Peters et al. (1988) do not draw conclusions specific for news language however, but merely treat the data as text samples representing Australian English.

The approaches in use and the perspectives from which Australian newspapers have been analysed are thus manifold and reflect the various interests in news language and content. A trend which is observable in the research, of which the above-mentioned studies are only a selection, is that apart from works that focus explicitly on community papers, certain newspapers are dominant as sources for data. Most common by far is *The Sydney Morning Herald*, which is also usually included when

Australian news is compared to news from elsewhere. Since so many studies take a qualitative approach, very often only two or three newspapers are used as a basis, whereas those that work quantitatively include a wide range of papers as they can draw on the huge variety created by the numerous regional publications. The importance of regional news should therefore be kept in mind during the analysis in the second part of this study.

2.3.3.2 Hong Kong

In terms of population relative to geographical size, Hong Kong is the extreme opposite of Australia. As was mentioned in section 2.1.3, Hong Kong covers an area of just 1,108 km^2 but has a population of roughly 7.3 million, resulting in an average of about 6,588 people per km^2 (HKCSD 2015, 4; CIA 2016). The literacy rate was 97% in 2011 (Ethnologue 2016), and in 2012 an estimated 99% of all households had access to radio and television (HKCSD 2012, 27). Furthermore 79.2% of all people were estimated to have regular internet access in 2014 (CIA 2016).

The media landscape in Hong Kong is of course closely linked to that of mainland China, and many newspapers that are sold in Hong Kong are imported from there. Nevertheless, due to Hong Kong's role as a business location and international region, a number of local newspapers have developed. Many of these are published in Chinese, but there are three major English-language newspapers in Hong Kong, the *South China Morning Post*, *The Standard* and the *China Daily Hong Kong Edition*. Of these three, *The Standard* has by far the highest circulation as it has been distributed free of charge since 2007 (The Standard 2015). Apart from these daily papers there are several English-speaking international papers that publish special Asia editions, most notably *The Wall Street Journal Asia* and the *Financial Times Asia*, and an online-only daily paper, the *Hong Kong Free Press*:

Table 2.2: Average circulation of English-language daily newspapers in Hong Kong, Mon-Sat editions

Newspaper	Circ.	Owner/Publisher
The Standard*	207,118	Sing Tao News Corp
South China Morning Post*	101,757	SCMP Group
The Wall Street Journal Asia*	81,348	News Corp
China Daily HK Edition*	40,000	China Daily Group
Financial Times Asia	31,784	Pearson
Hong Kong Free Press	n/a, online only	n/a, non-profit

(International Federation of Audit Bureaux of Circulations 2013; China Daily 2014; Financial Times 2014); *newspapers included in this study

The *Hong Kong Free Press* was founded in 2015, after the corpus for this study was finished, and is therefore not included; instead, *The Wall Street Journal Asia* completes the set of four papers for the Hong Kong component due to having a much higher circulation than the *Financial Times Asia*. All four papers belong to different news corporations and have been reported to reflect different political tendencies. However, in general external influences on the media and restrictions on reporting have seen a drastic increase in Hong Kong over the last years as the country was listed 34th in the World Press Freedom Index in 2010 and only 70th in 2015. In the first index in 2002, Hong Kong was ranked 18th (Reporters without Borders 2015).

The issue of press freedom is reflected in the research conducted with regard to mass media in Hong Kong. A number of studies (e.g. Lee 1998; Lee and Chan 2008; Chan and Lee 2011) make use of surveys among journalists and other media representatives to determine their attitudes towards press freedom and self-censorship, an aspect of reporting that became very frequent in Hong Kong after the reunion with mainland China. This is also the focus of a study by Lee and Lin (2006), which compares different Chinese newspapers from Hong Kong to determine how stories are written and framed in order to deal with political pressure, keep journalistic values and not lose the trust of their readership all at the same time. As both press freedom and self-censorship are closely intertwined with politics, other studies explicitly look at these phenomena regarding political events like the disputed independence of Taiwan (e.g. Cheung 2003) or, most prominently, the 1997 handover of Hong Kong (e.g. Sciutto 1996; Lee and Chu 1998; Lee 2000) to understand political influence and get an idea of Hong Kong's media development under these exceptional circumstances. As a sub-topic of the aspect of self-censorship, Yung (1996), Scollon (1997) and Scollon and Scollon (1997) look at the way sources are named, omitted or framed in Chinese and English-language newspapers from Hong Kong and mainland China.

This research interest into the aspect of press freedom and self-censorship is complemented by a number of studies dealing with the media landscape as such. Lee et al. (2002) analyse the logistics of media corporations during big political events, focusing on the handover of Hong Kong from Britain to China as an exemplary occasion, and Fung (2007) describes the facets of political attitudes that different newspapers can be observed to support. The other side of this coin, the readers' perception and reactions, is taken up for instance by Lee and Tse (1994), who analyse the change in media reception in immigrants from foreign cultural backgrounds with their arrival in Hong Kong, and by Flowerdew and Leong (2010) who, on the basis of articles from the Chinese-language paper *Ta Kung Pao*, try to determine the knowledge journalists presume their readers to have. Furthermore, Willnat and Zhu (1996) trace media priming in Hong Kong media by analysing the content of four papers and their coverage of an election campaign and relating them to the weekly polls concerning the election, showing that extensive reporting of an issue clearly influences the public's perception of its significance.

This particular finding was confirmed with regard to a different topic in a study by Yip et al. (2006), in which news items about the suicide of a celebrity were put into relation to an increase of suicides in general, demonstrating that the very elaborate and sensational nature of many reports can be seen to have an effect on the public. In this context, the study draws on the recommendations for reporting suicides as provided by the World Health Organisation (2000) and points out that many newspapers in Hong Kong are hardly in line with these. The study does not focus on linguistic aspects, but the very same point is made by two other studies which look more closely at the actual content of suicide reports: both Au et al. (2004) and Fu, Chan, and Yip (2011) use the official recommendations as starting points for their analyses and come to the conclusion that news reports on suicides are often selective and sensational, and rarely follow the recommendations for such cases. While both studies work with comparative approaches across different newspapers, they only use data from papers published in Chinese. The same holds true for another study into a socially relevant topic, namely the portrayal of women in official positions, which is examined qualitatively in Lee (2004). The most prominent examples of studies with a sociological focus based on English-language newspapers in Hong Kong are Chan (2014) and Cheng and Lam (2010) – the former analyses the creation of in-groups and out-groups in articles from the *China Daily* with regard to Chinese-Japanese relationships, employing psychological models for the discussion, while the latter uses data from the *South China Morning Post* to trace changes in the representation of human rights issues.

Apart from these, there are only few dominantly linguistic studies which explicitly focus on or include English-language newspapers. The earliest, Taylor (1989), is a very general work comparing the *South China Morning Post* and *The Standard* in terms of style and content. In contrast to this, other studies are significantly narrower in their foci; Carless (1995) examines the way Chinese idioms are translated and embedded in articles from the *South China Morning Post*, and Scollon (2000) compares English and Chinese editions of the *China Daily* and identifies differences in text categorisations between the two versions. Flowerdew, Li and Tran (2002) look at the *South China Morning Post* and compare discrimination markers in the language used by news and editorials.

Taking a more comparative perspective, Noël and Auwera (2015) analyse the usage of modal verbs in Hong Kong, the UK and the USA, by employing a quantitative approach based on data from *The New York Times*, *The Washington Post*, *The Guardian*, *The Times*, and the *South China Morning Post*. These three regions are also compared in Pan's (2002) study, which analyses the news coverage regarding the disputes concerning the Hong Kong right of abode in *The New York Times*, *The Times*, the *South China Morning Post* and the *China Daily*. Zhou (2008) looks at cultural differences in news items regarding internet usage in China, also drawing on data from the USA, the UK, and Hong Kong, and including Singapore as well. Hong Kong is represented in this study by one Chinese-language paper, the *Sing Tao Daily*, and the English-language *South China Morning Post*; however, Hong Kong-internal differences that could be traced back to multilingualism in the society are not addressed. Instead, it was shown that while the newspapers from the UK and the USA regularly address the issue of internet censorship in China, the papers from Hong Kong and Singapore put their emphasis on business matters and e-commerce, thereby avoiding any critical stance towards political decisions and demonstrating the self-censorship mentioned above.

The issue of self-censorship thus has to be regarded as a typical feature of Hong Kong media, and it is unparalleled among the press from the other four varieties included in this study. Although the analyses in this work do not claim to examine this characteristic in detail, it should be kept in mind when discussing the results and drawing conclusions, especially in the thematic domain of political news.

2.3.3.3 Kenya

As described in section 2.1.3, Kenya has a total size of 581,309 km^2 and its population is currently estimated to be 45.5 million, creating an average of 77 people per km^2. The biggest city in Kenya is its capital Nairobi with over 3.76 million inhabitants (United Nations Statistics Division 2016). The literacy rate in 2015 is estimated by UNESCO to be at 85.9% (UNESCO Institute for Statistics 2015).

In contrast to the other regions included in this study, the vast majority of Kenyan households does not have a television set. In 2010, only 30% were equipped with a TV, while 79% owned a radio set. Furthermore, only 7.9% of all households had a computer, which limited the internet access drastically (Nyabuga and Booker 2013, 14). In 2013, with the help of falling prices for mobile devices and respective contracts, the access rate had reached 31.7%, with almost 75% of Kenya's inhabitants having a mobile phone subscription (Kenya National Bureau of Statistics 2014, 56). Especially in rural areas, availability of television and internet is still very rare however (Kenya National Bureau of Statistics 2010, 423-26), a major problem being not the technical equipment alone but the access to electricity. The World Bank registers that only 23% of Kenyans had access to electricity in 2012 (The World Bank Group 2012), while The Open Society Foundation counts a mere 20% (Nyabuga and Booker 2013, 14).

With these figures in mind, it is not surprising that newspapers are a preferred medium for many people in Kenya (ibid., 18). The number of nationwide newspapers is fairly limited: there are five daily papers published in English, the *Daily Natio*n, *The Standard*, *The People*, *The Star* and the *Business Daily*, and just one paper published in Kiswahili, the *Taifa Leo*. Most of these papers also have weekend editions which are assumed to have a higher readership than the weekday ones. Two other dailies, the *Kenya Times* and the *Daily Metro*, stopped business in 2010 (Michieka 2005, 178; Nyabuga and Booker 2013, 20). In addition, there are three weekly papers, *The EastAfrican,* the *Citizen Express* and the *Coastweek*. The biggest media corporation is the Nation Media Group, which publishes the *Daily Nation*, the *Taifa Leo*, the *Business Daily* and *The EastAfrican* and also includes several TV and radio channels (Daily Nation 2016).

Determining the readership numbers of individual papers in Kenya is quite difficult; on the one hand, many people share newspapers and the actual audience can therefore be expected to be much higher than the amount of sold copies. On the other hand, there are no official figures from independent institutions regarding the circulation of papers, but only figures provided by the media corporations themselves and rough estimates from externals. In general, reports agree that the *Daily Nation* is the biggest newspaper not only in Kenya but in all of East Africa (Wachira 2014). As the *Taifa Leo* is published in Kiswahili and the *Citizen Express*, *The EastAfrican* and the

Coastweek are weekly papers, they were not included in the corpus for this study. Instead, apart from the *Daily Nation* the collection contains *The People*, *The Standard* and *The Star*, all four of which belong to different media corporations, as can be seen in Table 2.3:

Table 2.3: Average circulation of English-language daily newspapers in Kenya

Newspaper	Circ.	Owner/Publisher
Daily Nation*	170,000-180,000	Nation Media Group
The People*	110,000	Media Max Networks
The Standard*	48,000-60,000	Standard Group
The Star*	15,000-20,000	Radio Africa Group
Business Daily	8,000	Nation Media Group

(Nyabuga and Booker 2013; Cheeseman 2014; Wachira 2014); *newspapers included in this study

Despite this diversity in ownership, it should be kept in mind that political influence on media coverage is strong in Kenya, and Reporters without Borders lists the country on rank 100 of 180 in their Press Freedom Index for 2015, which is a huge decline from its 70th rank it still held in 2010 (Reporters without Borders 2015). Altogether, Kenya displays the highest degree of press freedom violation from the countries included in this study.

This instability concerning journalistic work, together with the still ongoing development of infrastructure which means that access to media is restricted for many people, leads to a lot of research into the Kenyan media landscape being primarily descriptive. A first account was given by Carter (1968), and several studies, such as Ochilo (1993), Skandera (2003), Djankov et al. (2011) or Ogola (2011), have since tracked changes and developments by including media descriptions in broader analyses of the country. These instances of 'taking stock' of what is there are complemented by studies that look at the role media plays in the society and the mechanisms that account for these roles: Ismail and Deane (2008) for instance analyse the positions and attitudes of different media throughout and after the country's general election in 2007, gathering data by conducting a survey of journalists and editors. Ong'a (2008) also uses a survey for journalists, determining which values and ideals young reporters hold about their profession.

Studies that look more closely at the content of newspapers are scarce, and as could be observed for the other varieties, are often driven by very specific, not necessarily linguistic, motivations. Amin et al. (2007) analyse the way newspapers report on malaria treatment, Machera (2001) concentrates on the representation of domestic violence, and Jacobson (2007) analyses advertisements for solar energy in Kenyan papers, thus taking biological, sociological and economic foci. Most of the linguistic studies taking Kenyan newspapers into account take comparative approaches: Schaefer (2004) and Hänsel and Deuber (2013) make use of quantitative methods, the former to look at framing in news stories about the 9/11 attacks in US and African newspapers, the latter to analyse Americanisms in newspapers from Kenya, Singapore and Trinidad and Tobago. Pollak et al. (2011) employ text mining on a collection of newspaper articles from Kenya, the USA and the UK to determine the depiction of the 2007 Kenyan elections in these different countries – however, the focus of this study is on the method rather than the actual news. In contrast to these studies, which draw on their own compilations of data, Schmied and Hudson-Ettle (1996) make use of the ICE component for East Africa and analyse *–ing*-constructions in newspapers from Kenya and Tanzania. A notable exception from the pattern of comparative research is a study conducted by Matu and Lubbe (2007), which combines the functional concept of transitivity with ideas from CDA in order to analyse the way in which political parties and their ideologies were constructed in newspaper editorials before Kenya's general election in 1997.

Although empirical studies with a focus on language are thus fairly rare, it is apparent that the *Daily Nation* is the most common source of data for such an approach. All studies described above include articles from this newspaper: the comparative works by Schaefer (2004) and Hänsel and Deuber (2013) draw exclusively on the *Daily Nation* to represent Kenya, and Matu and Lubbe (2007) and Pollak et al. (2011) combine it with articles from *The Standard* and the *Kenyan Times*. The ICE component used by Schmied and Hudson-Ettle (1996), although also featuring items from the *Kenyan Times* and *The Standard*, again mainly relies on the *Daily Nation*, and refers in its manual to the "dominance" the paper has in its country (Hudson-Ettle and Schmied 1999). In the study at hand, the *Daily Nation* is also included, but only contributes one fourth of the whole Kenyan component and will therefore show no more dominance in the results as the other papers included. Nevertheless, insights that have so far been gained into newspaper language in Kenya are often based on data collected only or predominantly from the *Daily Nation*, and cannot necessarily be generalised, a fact which should be kept in mind during the analysis and the discussion of the results later on.

2.3.3.4 UK

The United Kingdom is by far more densely populated than Kenya, but the density varies greatly across its parts. In total, the UK covers an area of 242,495 km^2 and in 2014 had 64.4 million inhabitants, reaching an average of roughly 266 people per km^2. Of these, 54.3 million lived in England (ca. 417 per km^2), 5.3 million in Scotland (ca. 69 per km^2), 3.1 million in Wales (ca. 149 per km^2) and 1.8 million in Northern Ireland (ca. 136 per km^2) (UK Office for National Statistics 2015b, 15). In 2011 Ethnologue estimated the literacy rate in the UK at above 99% (Ethnologue 2016).

In contrast to the other countries included in this study, the UK of course holds a particular position. To all four regions it is deeply connected historically, but the outcomes are very different. The USA and Australia were not primarily acquired as colonies for economic reasons but were intended as permanent settlements, which meant that not only people but also culture was imported to these parts of the world. The Englishes spoken in these countries are native varieties that were, in Kachru's terms, "transplanted" there (1981, 16-17; see section 2.1.2) and English became the dominant language almost naturally. In the cases of Kenya and Hong Kong, which were valued almost exclusively for economic reasons, the language was imposed on the country as far as necessary, and, as was seen in section 2.1.3, the status it holds there today is by no means undisputed. The UK is thus closer to Australia and the USA in terms of culture and the status of English, and the different colonial histories are not least reflected in the infrastructure displayed in these countries today. While Kenya, as was described in the previous section, still struggles with a comparatively low literacy rate of 85.9% and basic infrastructural issues like electricity supply, almost 86% of households in the UK had internet access in early 2015 and since 2002, 99% of all households have been registered as having a television and a telephone (UK Office for National Statistics 2013, 6; 2015a, 1).

However, with regard to mere geographical proportions the UK is nothing like the USA or Australia, a fact which is reflected in the media landscape. Of course there are numerous regional and local papers, but the major newspapers are all national ones and do not assign themselves to any particular region. In total there are 13 daily national papers, many of which publish extra editions on Sundays or weekends as well. The most widespread paper is *The Sun*, which in September 2015 was registered with 1,800,233 daily copies, followed by the *Daily Mail* with 1,609,003 copies (UK Audit Bureau of Circulations 2015, 2). Both of these are categorised by the UK Audit Bureau of Circulations as popular press (see also Conboy 2006, 7-9; Bednarek 2007, section 4.1) – the largest quality paper according to circulation numbers is *The Daily Telegraph*, followed by *The Times*, the *Financial Times*, the *i* and *The Guardian* (see Table 2.4). To ensure stylistic as well as ideological diversity in the analysis in this study, the corpus aimed at including both quality and popular papers and therefore contains *The Sun* and the *Daily Mail*, but not the third in the list, the *Daily Mirror*, which is also considered a popular paper (Conboy 2006, 7-9; Jaworska and Krishnamurthy 2012, 407). Instead, it includes *The Daily Telegraph* and *The Guardian* as representatives of the quality press (Gabrielatos and Baker 2008, 9). *The Guardian* was chosen since both *The Times* and the *Financial Times* are not freely accessible online and the *i* has a comparatively limited archive due to its recent founding in 2010.

Table 2.4: Average circulation of English-language daily newspapers in the UK, Mon-Sat editions

Newspaper	Circ.	Owner/Publisher
The Sun*	1,800,233	News UK
Daily Mail*	1,609,003	Associated Newspapers Ltd
Daily Mirror	855,698	Trinity Mirror plc
The Daily Telegraph*	480,873	Telegraph Media Group
Daily Express	418,001	Express Newspapers
Daily Star	403,380	Express Newspapers
The Times	392,572	News UK
i	277,498	Independent Print Ltd
Financial Times	206,813	Financial Times Ltd
Daily Record	186,439	Scottish Daily Record & Sunday Mail Ltd
The Guardian*	166,977	Guardian News and Media
The Independent	58,002	Independent Print Ltd
The Morning Star	25,000	People's Press Printing Society

(Dowell 2006; The Morning Star 2015; UK Audit Bureau of Circulations 2015); *newspapers included in this study

In contrast to the other four countries treated in this study, the UK has not displayed much variation in terms of press freedom since the establishment of the Press Freedom Index in 2002. Starting on rank 21 in the first index, its highest ever rank was 19 in 2010. Rather worryingly, the three lowest values are these of the last three years, declining steadily from rank 29 in 2013 to 33 in 2014 and finally to its lowest point yet, rank 34 in 2015 (Reporters without Borders 2015). In the current study it is still second only to Australia however, despite the negative trend, and the issue is therefore not a major focus of research into its media landscape. Rather, the question of objectivity and media democratisation arises in a number of studies, most obviously in the sense of means employed by journalists to incorporate their own opinions and attitudes into their writing (e.g. Chen 2005; Bednarek 2007). Other, related aspects are the attribution – or omission – of sources and potential ways of referring to actors included in a story (e.g. Jucker 1996; Smirnova 2012) as well as the degree to which user-generate content is allowed and integrated into online news (e.g. Hermida and Thurman 2008).

The question of objectivity, though not always explicitly expressed, resonates in other studies, especially those with an ideological focus. As was observable in the research concerning Australian news and will be even more visible regarding the USA (see section 2.3.3.5), one major topic in this direction is the representation of Islam and Muslims in the press (e.g. Richardson 2004; Baker 2010a; Baker, Gabrielatos and McEnery 2013). Especially in more recent years, this has often been connected with the public debate about asylum seekers and immigrants, and the interest in this is reflected both in individual studies (e.g. Matthews and Brown 2012) as well as bigger frameworks like the Lancaster-based RASIM project (Baker et al. 2008; Gabrielatos and Baker 2008; KhosraviNik 2008, 2009, 2010a, 2010b). The latter also allows resources for different approaches, and includes both qualitative and quantitative analyses building on the concepts of CDA.

The construction of “otherness” and “strangeness” that these studies have found to often underlie the representation of immigrants and asylum seekers is also present in reports on other countries, as an analysis conducted by Brookes (1995) into the depiction of Africa shows. It is opposed by the concept of national and regional identity, which can be constructed by discursive patterns (e.g. Bennett 2007), but also by the use of regional languages like Gaelic in Scottish papers, as shown by Cormack (1995). The aspect of identity is particularly important in articles concerning war, and the concept becomes more complex when the coverage switches from negative to positive. As for Australian and U.S. news, research into this field in the British press has focused strongly on the Iraq War; Murray et al. (2008) compare reports from before and after the war drawing on seven daily newspapers, Dirks (2006) analyses the representation of the USA in this context in the German and British quality press using *The Times*, *The Independent* and the *Frankfurter Allgemeine Zeitung*, and Rodríguez (2006) compares *The Sun* and *The Times* as representatives of popular and quality press in terms of their construction of the register parameters of field, mode and tenor of discourse. From a more anthropological perspective, Seymour (2004) traces differences and similarities between reports about modern Iraq on the one and ancient Mesopotamia on the other hand.

While these studies represent foci in the research into UK media, other political aspects have been analysed on a smaller scale in individual papers. Very often, they connect to issues mentioned before; for instance, Chen (2007) refers to the concept of objectivity when analysing attitudes towards politicians as expressed, or implied, by journalists from the UK and China. Taking up the issue of national identities, Wimmel (2009) and Breeze (2014) both discuss the representation of other countries, the former in the context of a potential inclusion of Turkey in the European Union, the latter by analysing the framing of Spain during the financial crisis. In an interesting study by Semino (2002), the mechanisms that usually apply to the representation of nationalities are examined in articles concerning the Euro. A more internal perspective is taken by Nessheim (2007, 2008), who analyses the press coverage on the British government during the time in office of Tony Blair, showing that British newspapers are by no means critical only towards non-British topics.

As for the other varieties in this study, there is a number of topics that has attracted researchers' attention apart from the ideological ones listed above. Two of these which can be found in Australian and U.S. related research as well are the representation of gender (e.g. Hawes and Thomas 1995; Kleinke 2000; Jaworska and Krishnamurthy 2012; Johnson and Ensslin 2013; Baker and Levon 2015) and the adaption of scientific issues for a broader audience (e.g. Koteyko, Brown and Crawford 2008; Camus 2009; Taylor 2010). Other studies treat less common areas, like the discourse on whaling (Murata 2007) or the reporting of homicides (Peelo et al. 2004), or look at very specific events and their news representation: Ermida (2009) traces the development of the coverage on the missing girl Madeleine McCann throughout a corpus drawing on *The Sun* and *The Times*, Bennett and Rowbottom (1998), using a qualitative approach, analyse the framing of Lady Diana after her death in 1997, and McEnery, McGlashan and Love (2015) compare the reports of 19 daily, weekend and online newspapers to the social media comments about a murder of a British soldier.

Although many of these works are linguistic in nature, they are inspired by sociological questions rather than by a desire to learn more about news language in particular. In contrast to this, a number of studies consider news without the implication of a specific sociological or ideological framework, and instead focus on either more news related subjects such as structure or type of news, or on specific linguistic features of news. One frequently found perspective concerning newspapers in the UK is a diachronic one, applied by studies describing either the development of media and news discourse (e.g. Brownlees 2006) or changes regarding the language and the content (Schneider 1999; Westin 2002; contributions in *Corpora* 5 (2), 2010[16]). In synchronic terms, Mahlberg (2009) for instance analyses the links between the text and the layout of a story, while Axelsson (1998) and Reynolds and Cascio (1999) analyse the use of contractions in British news writing. Other works on the lexico-grammatical level include an analysis of syntactic variation across 11 newspapers (Jucker 1992), collocations of proper names and nouns (Rydén 2008) and properties and meanings of very specific lexical items such as *fresh* (Mahlberg and O'Donnell 2008), *blue* (Szczygłowska 2013) and *problem* and *solution* (Scott 2001). Especially these last three rely on corpus-based approaches and draw on *The Guardian* and, in the case of Szczygłowska (2013), the *Economist*.

In general, no real tendency can be observed of one particular paper being overly dominant in the research; most studies include either *The Guardian* or *The Times*, but since quantitative, and in many cases corpus-based, approaches are most common, the analyses often contain three or more papers. A particularity of UK-based research is that it is often used in comparative contexts, either with non-English newspapers or in contrast to other varieties. This is also notable for research into newspapers from the USA, as the following section will show, and demonstrates that these two varieties still hold a privileged status as comparative values for English.

[16] Volume 5, Issue 2 of *Corpora* is a special edition entitled "Modern Diachronic Corpus-Assisted Discourse Studies" and contains several publications focusing on newspaper language from a historical or diachronic perspective. Papers include Clark (2010), Duguid (2010a, 2010b), Marchi (2010), Partington (2010) and Taylor (2010).

2.3.3.5 USA

Looking at the media landscape in the USA, certain similarities to Australia are undeniable. With a size of 9,629,091 km^2 it is the largest country included in this study, and has the second lowest population density with an average of 33 people per km^2 and roughly 322.2 million in total (U.S. Census Bureau 2015). Although it is thus still about 4.7 times more densely populated than Australia, there are areas with extremely low values, most notably Wyoming (2.3 people per km^2) and Montana (2.7 people per km^2). On the other hand, the biggest cities of the USA have extremely high density values, with New York City counting an average of 10,874 people, Los Angeles City 3,236 people and Chicago 4,614 people per km^2 (ibid.). The literacy rate in the country is estimated at being above 99%, and in 2014 86.8% of all households were estimated to have internet access (CIA 2016; Ethnologue 2016). 95.2% are currently registered to receive TV either via traditional broadcast, cable or internet (Nielsen Company 2015).

Due to the country's size, the newspaper landscape displays a trend towards regional rather than national papers, as could be observed in Australia as well. However, the USA has a much larger total population spread over more cities, which leads to more newspapers being published. All states of the USA have numerous local and regional papers, and the largest cities have their own metropolitan publications as well. The three largest papers are national ones, and many of the regional papers are published nationwide while at the same time keeping a more or less explicit focus on their respective areas:

Table 2.5: Average circulation of the top 15 English-language daily newspapers in the USA, Mon-Sat editions

Newspaper	Origin/Scope	Circ.	Owner/Publisher
The Wall Street Journal*	national	1,480,725	News Corp
USA Today*	national	1,424,406	Gannett
The New York Times	national	731,395	New York Times Company
Los Angeles Times	Los Angeles	432,873	Tribune Company
The Washington Post	Washington	431,149	Washington Post Company
Chicago Tribune	Chicago	368,145	Tribune Company
Daily News*	New York	360,459	Daily News
New York Post*	New York	299,950	News Corp
The Arizona Republic	Phoenix	285,927	Gannett

Newsday	Long Island	265,782	Cablevision
Tampa Bay Times	Tampa	241,020	Poynter Institute
Houston Chronicle	Houston	231,233	Hearst Corporation
Star Tribune	Minneapolis	227,694	Star Tribune Media Comp.
The Plain Dealer	Cleveland	216,122	Advance Publications
The Denver Post	Denver	213,830	MediaNews Group

(The State of the News Media 2012; Alliance for Audited Media 2013); *newspapers included in this study

Like for the selection representing the UK, we attempted to take into account both quality and popular press for the USA. The five biggest papers in terms of circulation are generally considered quality papers (Boykoff and Boykoff 2004, 127; 2007, 1194; Carpenter 2007, 766; Kleinnijenhuis et al. 2015, 7), but since neither *The New York Times* nor the *Los Angeles Times* were freely accessible, they were not included in the corpus. Instead, the more quality-end press is represented by the largest two in terms of circulation, namely *The Wall Street Journal* and *USA Today*. The next in line, the *Chicago Tribune*, again was not freely available; instead, the rather popularly oriented press is included via articles collected from the *Daily News* and the *New York Post*. Although both papers are located in New York, the *Daily News* is published all over the country and does not display an overly strong focus on its original area, and both papers feature national and international news.

Despite the apparent variation in ownership, the USA was registered with increasing problems concerning press freedom over the last years. Starting on rank 17 in the first World Press Freedom Index in 2002, it is currently ranked 49th (Reporters without Borders 2015). This is reflected in some of the research questions that have been investigated across U.S. newspapers; political tendencies of individual papers are analysed for instance in the works by Gentzkow and Shapiro (2006) and Larcinese, Puglisi and Snyder (2011), while the coverage of ideological and socio-political topics is dominant in general. Reports on environmental issues (e.g. Antilla 2005; Boykoff and Boykoff 2004, 2007; Bevitori 2010; Kleinnijenhuis et al. 2015) and political and foreign affairs (e.g. Lee and Craig 1992; Coutin and Chock 1996; Cotter 2011; Li 2011; Masumoto 2011; Greenberg and Miazhevich 2012) are the most common analyses; however, as could already be observed for newspapers in Australia and the UK, since the attacks on the World Trade Centre in 2001, a new research focus on aspects of terrorism has been observable (e.g. Yang 2003; Ryan 2004; Schaefer 2004). Furthermore, the outbreak of the Iraq War in 2003 triggered studies critically analysing its media coverage and framing in the USA (e.g. Dimitrova and Strömbäck 2005; King and Lester 2005; Carpenter 2007). Other societal issues such as reporting concerning Hurricane Katrina (e.g. Fahmy, Kelly, and Kim 2007; Potts, Bednarek, and Caple 2015), depiction of homo- and transsexual relationships and rights (e.g. McFarland 2011), the representation of gender norms and stereotypes in different news topics (e.g. Akhavan-Majid and Ramaprasad 1998; Meeks 2012; Aull and Brown 2013) or the shaping of identities of particular communities through the use of certain dialects or languages in local papers (Coupland, Bishop and Garrett 2003) constitute another major area.

Apart from these, more concrete topics are investigated from specialised perspectives, such as the depiction of tobacco use in newspapers (Clegg-Smith, Wakefield, and Edsall 2006), the use of diseases or ideological concepts as metaphors (Duckworth et al. 2003; Pérez-Sobrino 2013; Ming 2015) or the representation and popularisation of nanotechnology and other scientific topics in mass media (Pellechia 1997; Dudo, Dunwoody, and Scheufele 2011). These studies look at newspaper language from a non-linguistic angle and draw conclusions relevant for their respective fields. This holds true, to a certain degree, also for analyses conducted from the point of view of journalism studies, many of which look at media logistics or structures rather than at the language in use. Beam (1996) describes the market orientation of newspapers, and Benson et al. (2012) compare how the launch of online newspapers is conducted in the contexts of the different political situations in France, Sweden and the USA. Schultz (1999) and Barnhurst (2010) compare online and print newspapers and examine the differences with regard to layout and interactive potential.

From a linguistic perspective, several layers of news are interesting and have been analysed in newspapers from the USA. Rafail, Walker and McCarthy (2015) present a very general content analysis by determining which sorts of protests are reported on and which ones are usually left out or covered less extensively in articles from *The New York Times*. Ljung (2000) and Kornetzki (2012) go into more depth and focus on the text categories of U.S. newspapers in comparison to British and, in the latter case, also Russian ones. The papers by Danielson, Lasora and Im (1992) and Ljung (1997) work on a lexico-grammatical level, the former analysing readability of newspaper articles in comparison to novels, the latter comparing the text complexity across different news domains. Furthermore, very specific linguistic features are covered in the literature dealing with U.S. newspapers, such as the usage of proverbs (Lau

1996), modals (Ljung 1996) or the concepts attached to the word *strange* (Smotrova 2008). On a more abstract level, the aspects of evaluative language (e.g. Condit and Selzer 1985; Mei, Zhang and Yu 2015), the use and implication of sources and quotes (e.g. Makkonen-Craig 1999) and the expressed relationship between reader and author (e.g. Maddalena and Belmonte 2011) are frequent topics.

Just as the perspectives and disciplines which take an interest in the field of U.S. newspaper reports, the methods that are being applied display a wide variety. Linguistics is only one field out of many in this regard, and is mainly represented by sociolinguistic studies. Other fields include economics, psychology, sociology, communication science, politics and philosophy, to name but a few. Quite a number of studies take a multimodal approach, often focusing on pictures and the messages they convey (e.g. King and Lester 2005; Fahmy, Kelly, and Kim 2007). Others, such as Boykoff and Boykoff (2004) and Rafail, Walker, and McCarthy (2015) look at news from a diachronic perspective. In synchronic studies, comparative analyses of course offer themselves, and U.S. newspapers are being compared to both other English-language papers – especially from the UK – as well as foreign ones. In contrast to what could be observed for media analysis in Hong Kong, surveys and interviews are less frequent methods; in most cases, the method applied is predominantly quantitative, sometimes enriched with qualitative elements (e.g. Aull and Brown 2013; Kim 2014). For this, many studies compile their own collections to exactly meet their research interests, while others draw on existing corpora such as the newspaper parts of the Corpus of Contemporary American English (COCA), the American National Corpus (ANC) or LexisNexis Academic.

Independent of whether the data was collected specifically or was extracted from an already existing database, it is striking that almost all studies include *The New York Times*. As was seen above, it is the most widely circulated national daily newspaper in the USA and is thus dominant also in the research. If more than one U.S. newspaper is included, the *Los Angeles Times*, *The Washington Post* and *The Wall Street Journal* are most often used, followed by the *Chicago Tribune* and *USA Today*. Studies that are based on more than four or five newspapers usually include smaller, more regional publications, such as for example the *St. Louis Post-Dispatch* (Akhavan-Majid and Ramaprasad 1998), the *Tucson Citizen*, the *Arizona Daily Star* (Coutin and Chock 1996) or *The Buffalo News* (Duckworth et al. 2003). These papers are not included in contrastive works however; here, the national papers are the dominant choices, which is of course reasonable from a methodological point of view if the counterparts from the other countries are also supraregional papers.

2.4 Summary

The preceding sections aimed at giving an overview of the three areas that contribute to the theoretical foundation for this study, and to define and discuss key terms and concepts. The three fields can be seen as different perspectives on the topic of this work; variational linguistics and media studies provide the angles from which the data is to be considered, while the field of text linguistics offers approaches for the analysis and interpretation.

The combination of these areas offers very particular and detailed insights into both regional varieties of English and newspaper language. English is used very differently in the regions in which it is spoken, as its purposes vary and adapt to the needs of the respective society. Depending on the areas of life in which it is used and the people who use it most frequently, every regional variety has formed, and will continue to form, a characteristic set of registers. This register pattern thus reflects the status of English in that region and allows us to draw conclusions about its users and purposes. In order to understand it however, the history of English in that region has to be considered as it accounts for the official status of the language as well as the general attitude towards it. This historical dimension is reflected in the models described in section 2.1.1, however, all such models are restricted in their explanatory power; any attempt to define the different regional varieties by sorting them into categories can only give a rough overview, but leaves no room for the particularities of individual varieties or their specific linguistic developments.

The current study aims at looking at five varieties in more detail. While the UK holds a special position as the 'origin' of English, Australia and the USA are home to dominantly native varieties, and Kenya and Hong Kong to L2 varieties. Especially in the latter two, the status of English is fairly unclear though; although both recognise English as an official language and the respective varieties have been found to contain distinct features on various linguistic levels, it is difficult to determine the autonomy of the varieties and their societal status. A comparison with other varieties helps to put these Englishes into perspective and highlight their characteristic features.

Newspaper language is of course a very particular functional variety of language, and does not reflect every aspect of everyday life, yet it is a constant source of language relating to different topics and speakers, and can thus be regarded as a reasonable reflection of who uses English, and for which domains. Topics that are frequently discussed in English will have formed a distinct register, and whether or not a target group is expected to use English frequently can be expected to show in a newspaper's language. In countries with native varieties, these concerns are less relevant as a vast majority of readers is of course proficient in English, and it is by default the language to discuss almost every topic. In Kenya and Hong Kong, however, this is not necessarily the case, and newspaper language functions as a very good indicator for these matters. As was seen in section 2.3.3, many aspects of newspaper language have already been analysed in the different regions, yet studies that take a comparative approach focus either on regional or on functional variation. The former produce results about general differences between languages or varieties, but cannot inform about differences with regard to their functional diversity; the latter can come to conclusions about the degree of functional diversity within one variety, but by lack of comparative values cannot determine whether this diversity is characteristic for the respective variety or is a general tendency of English. A large-scale analysis into newspaper language in different regions which also takes into account functional variation within newspapers has so far not been conducted, but allows a combination of these perspectives and thus more detailed statements about the varieties in question.

Since the study, as will be described in the next chapter, contains a range of linguistic features which will be investigated for all domains and varieties, exact hypotheses for every analysis will not be formulated; based on the research that has so far been conducted into the media landscapes of the five regions as well as into newspaper language in general, a number of broad assumptions can be made however which will be tested in the analysis and suggest trends for the discourse under investigation. One of these of course relates to the distinction between native and L2 varieties; as was seen in section 2.1, English is not as dominant in Kenya and Hong Kong as in the other regions, and only selected newspapers are published in English. This in itself already implies that they address a specific target group, and that English is not the medium of everyday communication for everyone in the society.

This suggests that the news published in these two regions is likely to be more standardised, as the language is simply used less frequently and has therefore formed fewer registers as well as lower degrees of specialisation of the existing ones. Furthermore, a stricter application of journalistic norms would imply a higher level of formality and distance to create an impression of neutrality and objectivity. This is especially expectable for news from Hong Kong, on which self-censorship and the dependency on two different political contexts have been found to be a huge influence. This assumption is strengthened by the values regarding the Press Freedom Index, in which Kenya and Hong Kong are ranked lower than the other varieties and can therefore be expected to contain fewer instances of evaluative language and colloquialisms due to more severe censorship.

In more general terms, the different historical developments of the two regions have implications on their distance to the native varieties. Kenya's strive for independence from the UK was much more violent and problematic than Hong Kong's handover to China, which might lead Kenyan journalists to orient themselves towards the USA or Australia rather than the UK. Hong Kong on the other hand, as a result of its economic focus and growth as well as its political restrictions, can be expected to look towards those countries which are most important from an economic point of view, while constructing a more internal focus in political news.

Not only the L2 varieties can be expected to display particularities. A survey conducted by British Market Research Bureau Limited (Powell 2001) found that in the UK, roughly 20% of regular newspaper readers expected not only to be informed, but primarily to be entertained by their newspaper, and, as Harcup (2005, 89) states, "they are unlikely to be disappointed". Newspapers in the UK have proven to put a particular emphasis on entertainment (Franklin 1997, 7-10; Holland 1998, 31; Harcup and O'Neill 2001, 274), and can therefore be expected to show more deviation from journalistic norms and include linguistic markers characteristic of subjectivity and low distance between the discourse partners. The USA and Australia, too, are set apart from the other regions; as was already discussed in section 2.3.3, research has focused on as well as confirmed the importance of local news and publications due to the sheer sizes of the two countries. These tendencies can be expected to show also in the dataset at hand, although the relevance of regional and local events is likely to differ depending on the topic. For hard news, which deal mainly with accidents and crimes (see section 4.1.2), a regional focus appears logical as the events in question are more relevant for people close by, while for news from the areas of economy and politics, a wider picture, for instance on national level, is at least as important. For sports news, the focus can vary between regional and national scope depending on the event or team reported on, whereas for lifestyle articles, a tendency is hard to predict.

This particular domain has been the subject of surprisingly little research (see section 2.3.2), but is described in guidebooks for journalists as concentrating on human interest aspects or “soft news” (e.g. Harcup 2005, 116; Keeble 2006, 222; Boyd, Stewart and Alexander 2008, 22, 81). This is not to mean that the topics treated in this section are less serious, but that they are treated from a different angle and often at greater length (Keeble 2006, 219). For their linguistic realisation this suggests that the interpersonal dimension is constructed very strongly and the distance between the participants is kept low, for instance by means of more colloquial characteristics such as contracted word forms or sentence-initial conjunctions, but also by including more evaluative markers and direct addresses to the readers. As the norms that can be found for writing lifestyle news are less consistent and concrete, it can furthermore be assumed that this domain displays a fairly wide internal variation in terms of language use.

Independent of individual domains and varieties, a broader assumption can be made regarding the general differentiation. The five domains are, as will be described in detail in chapter 4 and was just indicated for lifestyle news, distinguishable on the basis of their topics and general orientations, which will of course be reflected in their linguistic realisations. For a comparison between the regions on the other hand, this will be less expressed as the domains are represented equally in the datasets for every variety, thereby neutralising the differences when subsumed under their respective region. While both domains and varieties are sure to display characteristics to some degree, the variation can thus be expected to be higher between the domains than between the varieties.

As functional variation depends on the actual language use and thereby the speakers and topics involved, the register framework in a systemic functional sense was found to be most applicable to this study. The framework mirrors these dimensions in the parameters it defines to describe a context of situation, namely the field, mode, and tenor of discourse. As was mentioned in section 2.2.2, these parameters are sufficient to cover a situation, but too abstract to analyse them linguistically. For this reason, each of the three has been defined to contain more clear-cut subdimensions which can be operationalised for an analysis on a textual level more easily. The following chapter will describe the operationalisation used in this study in detail and discuss the implications of a quantitative approach to newspaper discourse and register analysis to provide a background for its application in the second part of the study.

3 Defining Functional Variation

In principal, the dimensions which frameworks for functional variation offer are too abstract to be analysed on a linguistic level, and therefore need to be operationalised into concrete, observable indicators to guarantee a methodologically sound and comparable analysis. The aim of this chapter is to provide the operationalisation for this study and discuss linguistic features that have been used so far as well as their assignment to register parameters by different scholars. The chapter will therefore first give a brief description of the intended quantitative approach in general before turning to the three parameters of field, mode, and tenor of discourse and their respective subdimensions.

3.1 A Quantitative Approach to Variation

As was seen in the previous chapter, variation, both functional and regional, has been analysed using a number of different methods. For the analysis of regional variation, a qualitative approach is very useful to determine differences between varieties with regard to one linguistic feature as it can go into more depth in its examination, while a quantitative approach is advantageous if a number of features are taken into account. The same distinction of course holds true for functional variation, however, analysing individual features is less useful here in general. The language used by the participants of a communicative situation reflects all parameters of the situational context, in the case of the SFL-based framework it reflects field, tenor and mode. A thorough analysis of functional variation therefore has to take all of these into account, and an isolated linguistic indicator cannot mirror this complexity. Furthermore, in order to allow an adequate representation of functional variation, a certain number of texts has to be included in the analysis, or the results will have to be generalised too much (see section 4.1.2.3 for a discussion of representativeness). A quantitative approach offers the possibility to process large amounts of data and query it for several linguistic features comparatively fast.

Especially a corpus-based approach cannot be based on abstract indicators though; for automatic queries, the linguistic features have to be concrete. This in itself is not problematic for a register analysis, as the linguistic differences that mark functional variation can ultimately be found on the level of lexico-grammar (Halliday, McIntosh and Strevens 1964, 88; Halliday 1973, 51; Bartsch 2009, 105-6; Teich and Holtz 2009, 529) and thus offer themselves for systematic queries. More complicated than the actual search are the operationalisations, i.e. the definition of which linguistic features reflect and realise which register parameter.

In more detail, this will be discussed and defined in the following sections. What can be said already is that no parameter or subdimension is reflected by just one feature, nor by mere frequencies of several features. Instead, every register is "a cluster of associated features having a greater-than-random (or rather, greater than predicted by their unconditioned probabilities) tendency to co-occur" (Halliday and Martin 1993, 54). To determine this, statistical testing is inevitable, which again requires a certain amount of data and excludes the reasonable interpretation of functional variation on the basis of only one feature. In turn, a linguistic feature does not necessarily mirror just one parameter, but its use can have different implications which have to be considered in different clusters.

Despite these clear advantages of a quantitative, corpus-based approach to functional variation, it has to be noted that there are of course disadvantages as well. Some are purely technical in nature; every corpus has its limitations as to the complexity of linguistic features that can be queried, usually due to the limits in annotation and meta-information that the data includes. This leads to some potentially interesting features having to be left out as they cannot be queried. Furthermore, automatic annotators make mistakes and sometimes create systematic errors which might skew the results if unnoticed. Frequency counts and statistical evaluations, too, are never entirely accurate and of course inherit potential irregularities due to technical restrictions. Finally, every corpus-based study has to face issues like representativeness and data sampling for which there are various suggestions, but no ultimate solutions and therefore have to be considered with a great deal of care so as not to overgeneralise results or interpret findings wrongly. The detailed description of the dataset used for this particular study will be the object of chapter 4; during the following description of the register parameters and their operationalisations it should be kept in mind however that although the approach taken is mainly quantitative, qualitative elements will be included in some cases to prevent misinterpretations, for instance should a query produce extreme values for a specific variety or domain.

3.2 Register Parameters

The parameters that characterise the register definition in a systemic functional sense have already been mentioned several times in the previous sections and chapters. Functional variation can be defined along various parameters and on different levels of detail, however; while the SFL-based framework works with three dimensions, other approaches might use a different categorisation. Biber and Conrad (2009, 40; see also Fest 2011, 12) for instance define the situational context along seven parameters: participants, relations among participants, channel, topic, production circumstances, setting and communicative purposes. Another suggestion is made by Hymes (1967, 1974), who defines eight variables in his SPEAKING-model: setting, participants, ends, art characteristics, key, instrumentalities, norms, and genre. Both these examples work with a higher complexity on the first layer of parameters as they split the situational context into finer aspects than the SFL-based one, yet the tripartite approach contains all these aspects in its subdimensions and is therefore simpler but as effective while also mirroring different levels of abstraction (Fest 2011, 12).

The next sections will look at these levels and discuss the operationalisations that can do the respective parameters justice. In some cases, the nature of the texts used for this study is particularly relevant; in general, a broad operationalisation can be transferable to an analysis of an entirely different register and thereby allow comparisons of the results and thus of the different functional varieties. This implies that the operationalisation is not to take into account any linguistic features that might be considered (or have been found) to be specific of the discourse under investigation, in this case newspaper language, but is to be based exclusively on features that can be assumed to have the same or similar functions in every discourse. Although the comparability of the results is a great advantage, the focus of this work is on the differences between the varieties with regard to newspaper language, as well as between the news domains, and discourse-specific linguistic features might be of particular interest. Some such features will therefore be included where necessary or helpful, as the following sections will describe.

3.2.1 Field of Discourse

The first parameter that defines the situation is that of the field of discourse. In very general terms it has been defined as “the nature of the social action” or, in more colloquial words, “the general sense of what it is on about” (Halliday and Hasan 1985, 12, 24; cf. also Wegener 2011, 104). This does not only include the topic that is dominant in the communicative situation however; as Halliday (1974, 49) explains: “If, for example, the field of discourse is football, then no matter whether we are playing it or discussing it around a table we are likely to use certain linguistic forms which reflect the football context. But the two are essentially different kinds of activity and this is also reflected in the language”. Of course, it is not only reflected in the field of discourse but is observable in the other variables as well, yet here it is most dominant and obvious.

The subdimensions that the parameter of field is usually assigned are an attempt to capture this abstract concept of the “nature of the social action”. The topic is of course one of them, and is covered in the dimension of the experiential domain. It is complemented by a second variable which is to reflect the purpose the activity has in the authors’ eyes. Halliday and Matthiessen (2014, 37) call this “the nature of the social and semiotic activity”, which is confusing as the wording overlaps with the overall definition of field. What it implies is located on a lower level than field however; the authors distinguish between ‘doing’ and ‘meaning’ as the general functions language can have, and formulate seven more detailed ways of ‘meaning’ which essentially mirror the authors’ potential intentions. This subdimensions has therefore also been termed “goal orientation” (e.g. Neumann 2013) or “oriental goal” (e.g. Steiner 2004), which reduces the risk of misunderstandings.

It has been suggested to treat both the nature of the social activity and goal orientation as subdimensions next to experiential domain (see e.g. ibid., 14), however, the distinction between the two is not clear. Excluding the more concrete concept of goal from the variable of the nature of the activity leaves it very close, if not identical to the definition of the superordinate parameter of field and therefore renders it not only redundant, but also raises it again to a higher level of abstraction (cf. Neumann 2013, 48). An operationalisation that mirrors this would be very vague, which is why at this point, field will be discussed in terms of two subdimensions only, the experiential domain and goal orientation.

3.2.1.1 Experiential Domain

The essence of this subdimension of field, the experiential domain, is “to elicit the subject matter of the register” (Neumann 2013, 49; cf. Halliday and Matthiessen 2014, 33), that is to say it reflects the topic of the communication. There are two major approaches to determining this aspect, namely starting from a predefined categorisation of potential topics and sorting the texts in question into this pattern, or clustering the texts based on their similarity without any prior suggestions. The former, which is discussed for instance by Martin (1992, 544), has the advantage that texts are sorted not only in relation to each other, but into a bigger picture of potential fields of the language and society. Its disadvantage is that the defined categories have to be rather rough in order to cover everything, else the framework will get too fine-grained and lose its value as a basis for classification. In this study, the approach will not rely on any predefined pattern but will instead look at the differences between the varieties and news domains and let patterns emerge based on similarities.

In order to define the topic or subject matter of a text on a linguistic level, several indicators have been suggested throughout a number of studies. The most elaborate list is presented by Steiner (2004, 15), who names lexical fields, terminology, cohesive lexical chains, aspects of reference, headings and titles, paragraphing, transitivity, expressions of time, perspective and aktionsart. Neumann (2013, 49) makes use of the first four, condensing the features of terminology and lexical fields under “vocabulary” (cf. Fest, forthcoming[17]), which is mentioned by Halliday and Hasan (1985, 24) as well, and furthermore adds the operationalisation of lexical density in her analysis of field.

Not all of these features can be analysed in the dataset used in this study, as the texts include only limited meta-information and annotation. A proper analysis of lexical chains would require semantic annotation or a qualitative, manual evaluation of the data. Also the matter mentioned by Steiner of perspective is fairly abstract and difficult to determine on the level of lexico-grammar. Other features are less relevant in this study than in those for which they were originally operationalised; transitivity, for instance, is highly relevant for studies examining translations as a change from a transitive to an intransitive verb or vice versa in the translation process has an influence on the structure of the whole sentence and can trigger alterations with regard to

[17] Fest (forthcoming) is an earlier version of the analysis of the experiential domain in the sports part of the current corpus, examining keywords, tense and temporal expressions.

other linguistic features within field as well as the other parameters. This is not the case in this study, as it deals with varieties of the same language and does not include translations.

Furthermore, the interpretation of transitivity, no matter whether for one language or comparatively, is problematic. Halliday and Matthiessen (2014, 216) relate verbs to six "process types" which in turn feature up to three subtypes; for instance, "mental" process types include seeing, feeling and thinking, while "material" process types are instances of happening, changing / creating or doing / acting. These categories are partly overlapping and vary in degree of abstraction, which would make an assignment of individual verbs to these process types very subjective. Also, verbs can vary in their meaning, for instance when used in the context of an idiom or a metaphor, and a quantitative analysis would necessarily include a high error rate. Last but not least, when looking at Halliday's above-mentioned explanation of field on the basis of football, these process types, although part of the experiential metafunction, seem to be linked to field directly rather than via the intermediate step of the experiential domain. Verbs analysed here would serve as an operationalisation of the category of process type rather than of the experiential domain, and would allow conclusions about the "nature of the activity", but not necessarily about the topic. In this sense, Steiner's tripartite interpretation of field would be realised, but he, too, mentions transitivity as a feature of the experiential domain.

Due to these considerations and methodological concerns about increased subjectivity of the results, transitivity will not be part of this analysis. Instead, the experiential domain will be analysed in terms of vocabulary, expressions of time, use of tenses and lexical density. The vocabulary is the most straightforward realisation of the subject matter, as the mere frequency of a word is already an indicator towards the dominant topic. To avoid false conclusions, keyword lists, generated for each domain and each variety, will complement this category. Lexical density gives more general hints at the topic by measuring the information density of the text.

Steiner's mentioning of expressions of time is less straightforward than these operationalisations. Temporal references in themselves of course help to determine the scope of the subject matter, yet he gives no detailed explanation as to why he chooses to include expressions of time, but not of place. For the study at hand however this distinction is useful, as the two have different implications for newspaper readers.

Both recency, constructed by temporal references, and geographical closeness, constructed by spatial references, are aspects that make a story newsworthy (Bednarek and Caple 2014, 155; Potts, Bednarek and Caple 2015, 3-4), yet the former is less patronising as well as more limited in this context. News is by definition not old, so the span of time that is being covered is short and differences between individual items have to be determined on a very fine-grained level. The importance of recency can vary depending on the topic however; match reports or business news for instance expire faster than health-related stories or travel reports (see also Fest, forthcoming). Independent of urgency, the temporal aspect does not suggest in- and out-groups to the readers, as the same references of time hold true for everyone. The recency of a news story can therefore be regarded as part of the topic. Expressions of place on the other hand do not just construct the news value of geographical closeness, but also create the impression of a shared background between author and reader. Furthermore, if contrasted to places that are geographically far, in- and out-groups are suggested which reflect the target groups of newspapers. Place references are of course also topic-related, yet in a newspaper context they most of all mirror the importance of the respective place and its people and culture. For this reason, place references will be included in the tenor of discourse rather than here.

Recency is not the only temporal aspect that gives an indication about the topic though. While in general, news articles refer to events that have already happened, they can also include outlooks on future happenings or formulate speculations and for these purposes vary the use of tense and aspect. Although this does not give insights into the topic as concretely as for instance keywords do, in combination with the other operationalisations of this subdimension it can produce results as to which topics contain temporal variation and to which degree. This investigation will therefore include an analysis of tenses and *will-* and *going to*-future constructions.

In general, it should be kept in mind that a text can of course have more than one topic. The analysis of the experiential domain will show the dominant topics, yet the examination of the vocabulary will give some insights into the scope of the texts as well as the domains. The domains of course are to a certain degree a predefinition of topics; since the domains were not changed however but taken over from the newspapers, and texts were assigned not based on their content but based on how the newspapers categorised them, an analysis of the actual content remains an interesting aspect to analyse. Domains often overlap in the topics they cover: sports articles can contain economic content, crime can be associated with politics and lifestyle items can in turn be about sports, to name but a few examples. To determine differences in the newspaper language across the varieties and to define the degree of differences between the domains, an analysis of the experiential domain is therefore very useful.

3.2.1.2 Goal Orientation

The second subdimension that is considered for the field of discourse is that of the goal of the interaction. If language is treated as purposeful action, as was described in section 2.2.1, then it logically follows that the participants pursue a goal. A speaker's intention can be obvious or hidden to his communicative partners (see section 2.2.1, speech acts), but as Hasan (1999, 234) points out, the speaker's own awareness of their goals also varies depending on the situation, thereby making the purposefulness of the action somewhat scalable rather than absolute. In contexts with a focus on interpersonal, relational aspects, the awareness of particular goals has been shown to be lower than in action-based contexts (ibid.), where the achievement of a goal is immediate. To better define this difference, Hasan (ibid., 234-35) distinguishes between short term and long term goals and states that while short term goals, such as buying or selling something, are rather obvious, long term goals, like making friends or social contacts, are usually less visible and often also less dominant in the speaker's immediate consciousness.

In this sense, communication not only has different types of goals, but can also have more than one goal as well as feature combinations of visible and invisible and short term and long term goals. Halliday and Matthiessen (2014, 41) map potential goals onto the type of activity and its orientation, saying that communication can be oriented towards field or tenor or sometimes both. Goal types such as "expounding" or "doing" focus on the subject, while actions like "recommending" or "exploring" look more towards the interpersonal relation. Again, the boundary between the two is blurry, and most texts will feature both to varying degrees.

When focussing only on written texts, as is the case in this study, quite a number of the above-mentioned considerations become less relevant. First of all, the communication is one-sided, which means that only the author expresses any goals at all. The addressees, in this case the newspaper readers, are passive; this is not to say that they do not have goals, however there is no possibility to trace them linguistically as they do not produce any text in the communicative situation. Furthermore, producing a written text requires more awareness of one's intentions than some forms of spoken interaction, like small talk. In the case of journalistic texts, this is strengthened even further by the editing process articles go through and the numerous people involved in it.

In general, there are several suggestions of goal taxonomies for written texts which would serve an operationalisation. Engel (1988, 118-19; my translation) presents a model of six potential functions, namely to inform, to induce, to convince / persuade, to teach / correct, to maintain contacts, and to reduce emphasis. Brinker ([1985] 2005, 113-30; my translation) only mentions five functions, information, contact, declaration, obligation and appeal. Neumann (2013, 55-56) works with four, narration, argumentation, exposition and instruction. These different categories already show that the concept is not treated alike by everybody; while the former two by Brinker and Engels describe desired outcomes of texts (see also Halliday and Matthiessen 2014, 41), the latter by Neumann is more closely linked to the way these outcomes are achieved and the processes involved.

A more discourse-specific distinction is used by Lüger (1995, 66-74), who adapts his categorisation to media texts in particular. He defines five *Textklassen* which correlate which specific intentions, namely texts that are informative ("informationsbetont"), evaluative ("meinungsbetont"), appellative ("auffordernd"), instructive ("instruierend-anweisend") and contact oriented ("kontaktorientiert"). For each of these, he further determines different types ("Textsorten") of news, such as hard news, soft news and reports as types of informative texts, or comments and reviews as types of evaluative ones. These types are different realisations of their respective *Textklasse*, and are characterised by conventionalised structures and features.

Lüger's approach is mainly language-external; although he does define some linguistic characteristics for the different types, the categorisation of texts depends on their domain. According to his framework, the vast majority of texts included in this study are therefore informative texts, with the only potentially questionable cases occurring in the lifestyle domain which in some papers includes reviews, a type that Lüger (ibid., 139) classifies as evaluative, or *how-to*-articles, which would count as instructive. However, a top-down sorting of texts in this way leaves out the question of the linguistic realisation of the different goals. Since the articles for this corpus were taken over from the domains in which the newspapers themselves situated them (see section 4.1.2), a categorisation strictly according to Lüger's system would be very simple, yet it would not allow any detailed conclusions about linguistic similarity.

Neumann's (2013) categorisation would be more effective in this way as the categories are independent of news discourse, but still general enough to be applicable and make the results comparable. On the downside, this breadth and generality means that the categories are rather rough and are only of limited use to determine fine differences between texts from the same discourse area. Particularly an analysis on text level, which would reflect mainly the dominant function of each text, runs the risk of resulting in one big cluster instead of reflecting potentially small, yet interesting differences. A thorough analysis and weighing of different intentions and goals would have to look at paragraphs and sentences and take into account the goal-related structures within the articles. Since every text can be assumed to have more than one goal (see Lavid 1995, 33; Neumann 2013, 55), it would be the composition rather than the dominant function which would be of use for an analysis within one discourse.

Although the corpus does not contain mark-up for syntactic functions or parsing and an analysis within the texts is therefore difficult, goal orientation will still be included in this investigation in a simplified manner. In general terms, news texts would be expected to be, to varying degrees, informative in nature (Fest 2015, 54) – this goal can therefore be assumed to be present, at least marginally, in all texts in this corpus. How dominant this goal is, depends on a number of factors, including the topic, the assumed audience, the newspaper, and the presence and strength of other potential intentions. Keeping in mind that a text can have different goals and that concentrating on one aspect of course allows no conclusions about goals in newspaper texts in general, the analysis of goal orientation in this work will focus on the degree of informativeness the individual texts display to capture differences along this dimension.

While Engel and Brinker explicitly name information as a goal or function, they both agree that this goal is compatible with different means of realising it, e.g. by evaluative or factual presentation of the content (Engel 1988, 118-19; Brinker [1985] 2005, 115). For newspaper articles, which both address as examples, factual language is described as dominant (Brinker [1985] 2005, 116), which in Neumann's (2013) categorisation is most closely linked to exposition. The notion of exposition draws on the definition Beaugrande and Dressler (1981, 190) offer for descriptive texts, which focuses on “objects and situations” as the centre of this text type. Based on this assumption, the language can therefore be expected to be very factual in the sense that it is rather impersonal and content-oriented.

To determine the degree to which texts focus on facts, Neumann (2013), drawing on Werlich (1976), defines a number of indicators, some of which however have to be regarded critically for the case of newspaper discourse. One such indicator is the frequent use of present tense in informative texts; news by definition reports on something that has already happened, and sometimes includes outlooks on what might or is scheduled to happen in the future, but the respective tenses do not imply that the information density is lower. Rather, they reflect the news values which are dominantly constructed for the story. In turn, if an article is written in present tense, it does not necessarily follow that it concentrates more on information, but can also mean that this text discusses a topic which is less bound to a concrete event and time, like health issues or beauty tips. The distribution of tenses can therefore not be linked to the degree of informativeness without risking to draw false conclusions, and will not be used as an operationalisation here.

A similar problem arises with the suggested indicator of pronominal use. Neumann (2013, 58) argues that out of all personal pronouns, the singular third person *it* stands out as it is more factual than the others and therefore refers to facts more than to people. In this sense, *it* can be expected to occur much more frequently than other personal pronouns in more informative texts, while personal references in general are rare. This observation is perfectly reasonable for many written texts, yet in newspaper language, the mere inclusion of a person in a story can be the dominant news value (Bednarek and Caple 2014, 156; Potts, Bednarek and Caple 2015, 3-4) and thereby be an important, if not sometimes the most important, piece of information to give. In such a case, personal pronouns are to be expected just as much as *it*, without changing the information value. Rather, the frequency of all 3rd person pronouns can be taken as a measurement.

This value, too, brings problems however and needs to be modified further. Some prescriptive journalism books (e.g. Keeble 2006, 113; Papper 2013, 49-56) contain examples of using reference points to news actors in the text as slots for information and instead of using pronouns include facts, as the following example shows:

[1] **Heather Watson** is hoping glandular fever will not rule her out of next month's French Open. [...] **The 20-year-old** has not picked up a racket since her diagnosis

> three weeks ago and says she has no idea when she will be well enough to return to the practice court. (Hart 2013, *The Daily Telegraph*; emphasis added)

A low frequency of 3rd person pronouns in general can therefore be expected in news as journalists are trained to avoid them and their mere frequencies would be over interpreted without taking this into account. Instead, as an operationalisation for goal orientation the amount of 3rd person pronouns in relation to all pronouns will be analysed, as 1st and 2nd person pronouns reflect a more interpersonal dimension.

The potential relevance of actors also has an influence on the suggested operationalisation of frequent nominalisations. The decision to use a noun instead of a verb shows that the focus is on the event rather than on the process expressed by the respective verb and the actor involved. Yet in cases where the dominant news value is constituted by a person (or several persons), their respective involvements are a necessity and a preference for verbs cannot be interpreted as a reflection of a lower degree of informativeness.

Other suggestions by Neumann (2013) and also Biber (1995) are applicable to news language without alteration, most of all the dominant use of declarative mood which, in contrast to imperatives and interrogatives, puts a clearer emphasis on passing on information, and a high type-token-ratio, which signals a wider focus of information and reflects the use of different words to describe an event. Both of these features will therefore be included in the analysis.

3.2.2 Tenor of Discourse

The second parameter to describe the context of a communicative situation is that of the tenor of discourse. It is, in broad terms, concerned with all interpersonal aspects that might have an influence on the language in use and will contribute the largest parts to the analysis in this study.

Concerning its subdimensions and the respective terminology, there is a certain variation in the literature dealing with this parameter. In general, there are three aspects that are mentioned frequently, namely the agentive roles the participants take as well as the power relations and the social distance between them (Halliday and Hasan 1985, 57; Poynton 1989, 76-77; Hasan 1999, 233; Steiner 2004, 17; Lukin 2010, 100-1; Lukin et al. 2011, 199-201; Neumann 2013, 61-32). Another dimension which is sometimes added is that of affect or appraisal (cf. e.g. Martin 1992, 533-37; Neumann 2013, 61), which represents instances of emotion and evaluation expressed by the participants. In this context, Martin defines three scales along which the kind of evaluation can be determined, *positive – negative*, *self-oriented – other-oriented* and *predisposition – surge*. The Appraisal framework presented by White also contains three types of evaluation, namely attitude, engagement and graduation, which again disperse into several subcategories each (Martin and White 2005, 38).

While the first three dimensions of social distance, social hierarchy and agentive roles are located outside the language and have to be operationalised to be traced linguistically, evaluation appears to be on a more concrete level already. Despite the indisputably existing different kinds of evaluative language and the usefulness of the fine-grained models offered by the notions of affect and Appraisal, these are distinctions of linguistic characteristics on the level of lexico-grammar rather than on the context- or discourse-level. Since the expression of evaluation has often been found to be closely connected to the building of solidarity (e.g. Martin 2004; Drasovean and Tagg 2015), a notion which in turn is related to the subdimension of social distance (Martin 1992, 525; Halliday and Matthiessen 2014, 35), the following sections will work with three subdimensions and include evaluative language as one aspect of social distance.

3.2.2.1 Agentive Roles

The subdimension of agentive roles defines "references to semiotic roles assigned through the text to author and reader/ hearer" (Steiner 2004, 17). Every communicative situation requires certain roles to be filled, therefore the nature of the activity predefines the roles of the participants (Halliday and Hasan 1985, 57) which are then reflected in the language in use. Agentive roles do not necessarily stay constant during communication, but can be transferred from one participant to another as control of the interaction is shifting (Neumann 2013, 63). In the case of many written texts however, the addressee is virtual and producer and recipient are separated by both space and time (cf. Hasan 1999, 238). For newspaper articles this is certainly true; journalists write for an anonymous audience of which they have a profile in mind, but which they can never clearly identify. The agentive roles of the participants are therefore maximally static and cannot shift as there is no direct interaction. By definition newspapers provide information which the readers absorb, which necessarily leaves the author of the text in charge of the interaction and the reader in a passive state. As any change or shift of agentive roles is therefore simply not possible, an analysis of this subdimension is irrelevant for newspaper texts and will not be included.

3.2.2.2 Social Role Relationship

The second subdimension of tenor to be discussed here is that of the relationship between the social roles taken by the communicative partners, also sometimes referred to as "social hierarchy" (e.g. Halliday and Hasan 1985; Hasan 2009a, 2009b; Bowcher 1999; Steiner 2004; Scott 2011). In general, it reflects the power structures that are at work between the participants of a communicative situation (see e.g. Steiner 2004, 17; Neumann 2013, 63). When looked at in more detail, these power relations can of course be traced back to various aspects of both natural and societal origin. On the one hand, characteristics like gender, race, and age contribute to the hierarchical structures between people, although they are of course beyond their control and are therefore often referred to as "ascribed roles" (Johnson 1966, 140). They are complemented by "achieved roles" which are more 'artificial' aspects that can create power differences. Although these two types condition each other, for the authors included in the corpus of this study facts like age or race are not known in all cases, which is why these dimensions will not be considered further at this point; instead, the analysis will focus on achieved roles.

Halliday and Hasan (1985, 57) explain that "the degree of control (or power) one participant is able to exercise over the other(s)" is basically predefined by the agentive roles they hold. To a certain degree, this is true, and newspaper texts are no exception; the mere fact that readers read a newspaper to get information from the journalists already determines that in terms of expertise and authority, these journalists have to be superior. Yet despite the fact that the agentive roles, as was said in the previous section, are static in a newspaper context, the level of hierarchy which the journalist constructs can vary. Superiority, or at least equality, is a necessity in this communicative situation, but to what degree and along which lines is open and depends not so much on the agentive roles as on the field and the social distance between the participants. Also, it is the journalists' choice how to express this superiority linguistically, a decision which in turn reflects the knowledge the writers assume in their readers and the social roles they construct for their imaginary audience.

As reflections of the social role relationship, Poynton (1989, 76-77; emphasis in original) defines four aspects:

> **Force** involves physical superiority; **authority** is a function of socially-legitimated inherently unequal role relationships such as parent-child, teacher-child, employer-employee, or ruler-ruled; **status** is a matter of relative ranking with respect to some unevenly distributed but socially-desirable object or standing or achievement, e.g. wealth, profession/occupation, level of education, hereditary status, location of residence, overseas travel; **expertise** is a matter of the extent to which an individual possesses knowledge or skill, e.g. the expert knitter compared with the novice, the nuclear physicist with the high school student beginning to study physics.

Neumann (2013, 63-67) leaves out the aspect of force in her study, which is logical for an analysis of written texts, and works with the three dimensions of authority, expertise and education. Although Poynton mentions education as an example of status, Neumann (ibid., 65) assigns status in general to the category of authority, singling out education separately.

Although these aspects mirror different dimensions of a social role, on a lexico-grammatical level they are hard to tell apart as they overlap and linguistic characteristics therefore cannot be clearly assigned to any one of them. Especially education and expertise can be expected to find their way into a text via similar linguistic means, which makes a distinction at this point superfluous. Regarding expertise and authority, various suggestions have been made in the literature with regard to their linguistic representations; concerning authority, Poynton (1989, 8, 75, 81) suggests mood, modality and reciprocity of voice, the latter by pointing out that superiors in a communicative situation have the "rights to interrupt, to nominate topics, to be definite, etc. that subordinates do not have", in other words to control the communication and the direction it takes. Since in the case of this study, as in most written texts, there is no actual discussion that involves active participation from author and reader, reciprocity is not relevant here. Furthermore, reciprocity is, at least partly, located on the discourse level, while mood and modality are more clearly traceable on the level of lexico-grammar.

At first glance, mood seems the more obvious feature of the two, since especially imperatives are a clear indicator of who is in charge. Interrogatives, however, can be interpreted in different ways, either as a signal for equality by asking for the opinion of the communicative partner – and thereby recognising his expertise – or as a way of phrasing an imperative in a less direct way, which might be preferable particularly for subordinate participants (Poynton 1989, 71; Neumann 2013, 65). In the case of newspaper writing, either is possible, but due to the one-sided character of the situation and the discrepancy in time and space between production and reception, it is unlikely that an interrogative will trigger immediate action in a reader. If action is required, for instance in the form of calls to contact the police about information on a crime, or to avoid a motorway because it has been closed, this will be done via imperatives as it can be assumed that readers want such pieces of information in a straightforward way. Rather, the use of interrogatives can result first and foremost in the readers thinking about and possibly challenging their beliefs and opinions, thereby resembling the function of interrogatives in literature (cf. Krings 2003, 151; Fest 2015, 59; Wenzel 2015, 28), or work as an introduction to a topic and a teaser for the next pieces of information. Still, whether they increase authority or rather reduce it in this context depends on the individual cases, and the respective analysis will draw on qualitative spot tests to determine the dominant functions.

Modality, too, is not undisputed as a marker of authority. Poynton (1989, 71, 79) refers to the effect of modalised formulations to weaken a statement and thereby make it less definite, which can be seen as a characteristic of inferiority. However, as Neumann (2013, 66) points out, modality can signal other things as well and a discrepancy in terms of authority is not the only possible explanation. The use of modal verbs can for instance signal uncertainty about a topic or piece of information, which would then relate it to the matter of expertise rather than authority. As was already said in section 3.1 and was just described for mood as well, this multi-functionality of linguistic features is normal; in the interpretation of the results it is therefore important to keep this in mind and look at the occurrences of features not in isolation but in the contexts of the respective subdimensions they represent. In this light, modality will be included in the operationalisation for social roles and their relations alongside mood.

In addition to these two linguistic characteristics, there are other means that have been suggested to represent the aspect of expertise. Neumann (2013, 66) includes the use of technical terminology, arguing that this reflects language for specific purposes which by definition requires expertise. Other scholars, like Eggins and Martin (1997, 233), categorise this linguistic marker as an aspect of the field of discourse, which again shows the potential multiple functions of a feature. Independent of its categorisation, the aspect of technical terminology is hard to define. As the analysis will show, every domain is set apart by certain keywords which are more relevant there than in the other news, like *match* and *season* in sports news or *bank* and *stocks* in economy. This does not necessarily mark them as technical terms however, and a decision as to where to draw the line between common and technical is highly subjective and unconsciously dependent on the researcher's own expertise. As prescriptive journalism guidebooks (e.g. Pape and Featherstone 2005, 27; Keeble 2006, 99) advise authors to avoid too complex expressions altogether, technical terminology will not be defined more closely in this study. Instead, two other markers will be included as representations of specialisation on a more general scale. First, the amount of keywords will be examined (Neumann 2013, 144-45) to see how specialised the language in a certain domain or variety is in comparison to the other parts. Second, Bartsch (2009, 115; see also Biber 1988, 73-75) suggests the mean word length of texts as an indicator for specialisation, as technical terms in general tend to be longer than more frequent vocabulary.

The subdimension of social role relationship will thus be represented by the linguistic characteristics of mood, modality, amounts of keywords and mean word length. As was seen above, an exclusive assignment of a linguistic feature to either expertise or authority is very difficult; therefore, results should not be interpreted separately but always regarded as different perspectives on the dimension of social role relationship.

When discussing hierarchical structures in the analysis, it should be kept in mind that the relationship between reader and author in the case of newspaper articles is of course characterised by an entirely one-sided form of communication, and that any instances of hierarchy are constructed by the author alone. The reader cannot immediately respond – if the newspaper conveys a degree of hierarchy that is unacceptable for its audience it will probably notice this mistake most drastically in receding sales figures. For this reason, as Neumann (2013, 64) argues, the term *hierarchy* is problematic with regard to written texts as it describes a relationship between two known points. The concept of "social role relationship", which has been used as an alternative name for the category of social hierarchy (Steiner 2004, Neumann 2013), is more neutral in the sense that it more clearly includes the option of equality. Furthermore, in sociology social roles are defined both as realisations of norms that we as members of a society learn and as reflections of expectations others have of us (Schäfers 2002, 33; Scherr 2002, 52), and are therefore not fixed. In a one-sided communicative setting, where we can only analyse the producer's linguistic choices and draw conclusions on that basis, this aspect of expectations towards the addressee is more appropriate that to speak of the more objective notion of hierarchy. Another term that has been suggested for this subdimension is that of "status" (Martin 1992), which has the advantage of being the most neutral one both in denotative terms as well as regarding strong connotations in any particular scientific area. It subsumes both achieved and ascribed roles, yet it does not reflect at all the aspect of relating two or more participants, which is the essence of the parameter of tenor. Although all these terms are in their ways appropriate and useful, the present study will therefore work with the concept of social role relationship.

3.2.2.3 Social Distance

The subdimension of social distance is by far the one for which most potential operationalisations have been suggested in other studies. It describes first of all "the frequency and the range of previous interaction" (Hasan 1978, 231) between the participants. As Lukin (2010, 101) rightly points out, where the addressee is imaginary as is the case in newspaper writing, social distance in this sense must be assumed to be maximal, as we can simply not say whether the communicative partners have met before and if so, in what context. This side of the subdimension is reflected in the term "contact", which Martin (1992) and Poynton (1989) use instead of distance. However, social distance can relate to other aspects of a shared background as well, like growing up in the same city, attending the same school, or belonging to the same sports club.

In all of these cases, the communicative partners can be assigned to an in-group in a particular aspect, in contrast to an out-group consisting of all those who do not share this background. This side is reflected in the term "solidarity", which is used by Brown and Gilman (1960) and has been linked to the notion of social distance by several scholars (e.g. Martin 1992, 523-525; Martin and White 2005, 30, 35; Halliday and Matthiessen 2014, 35). Steiner (2004, 18) defines social distance as "the degree to which the contextual space is shared by participants", which mirrors both sides. As for the social role relationship, in newspaper articles this shared background is of course constructed solely by the journalist and we cannot know the reaction of the readers. Yet different forms of shared background are reflected in the criteria for news worthiness, for instance in the fact that events happening in places that are geographically or culturally near to the target group of the news have a higher priority (Fowler 1991, 13-14; Harcup 2005, 30; Kolodzy 2006, 60-61; Bednarek and Caple 2014, 155-56). We can thus assume that the construction of solidarity and a certain closeness to the readers is present in news, and via this construction, in- and out-groups as well as the assumed target group can be traced linguistically.

Concerning this linguistic realisation of social distance, Martin (1992, 531) first of all distinguishes between "involved" and "uninvolved" contact, depending on the degree of familiarity between the discourse partners and the frequency of their meetings. As concrete linguistic markers, he suggests a number of features including tone, accent, vocation, specialised terminology, ellipsis, contractions, use of names (e.g. nick-names versus full names), and slang. Many of these are relevant only for spoken language and can thus be left aside here. The use of names, too, is of limited consequence for a newspaper context with regard to the distance between author and reader. Nevertheless, the use of first or last names, titles and addresses does have a linguistic implication, as it can serve as a tool for the author to create a distance or closeness to the actors in the story (Fest 2011, 60, 62, 72; forthcoming) It makes a difference whether for instance the football player is referred to as "Mr Rooney" or just "Rooney", since the former appears strange for an athlete and hints at a potential seriousness of the topic. For politicians on the other hand, titles and addresses are more usual, and by leaving them out the author can signal disagreement or even disrespect and distance themselves from statements these people made.

Especially names are not easy to query in quantitative data however, as the part of speech-tagger does not differentiate between names of places or companies and names of people, let alone between first and last names or even nicknames. Particularly Asian and African names caused further problems as these are not standard items for a tagger trained on English and were therefore often tagged wrongly, which would skew the results when relying on the NP-tag (proper noun). With five different regions included in the corpus, the list of personal names is nearly infinite, in contrast to titles and forms of address, which constitute more closed classes. The analysis here will therefore focus on the latter two and regard them as an instance of evaluative language, which, among other effects, can create solidarity between participants (cf. Fowler 1991, 63).

This is not the only effect of evaluation, of course; Martin and White (2005, 30) assign it to the subdimension of status or social role relationship rather than to that of social distance, arguing that the use of evaluative language reflects "who can express feelings and who can't, what kinds of feelings are expressed, how strongly they are expressed, and how directly they are sourced". In a direct communicative situation this is very helpful as an indicator for hierarchy, however, since in the present study the communication is one-sided, the presence or absence of evaluative language cannot be taken to mean that the respective journalist can or cannot afford to express emotion in comparison to the audience. The concept of affordance implies that who can utter evaluation is in a position where he or she does not have to seek agreement or approval from others, but has the status to express an attitude rather freely. Since a newspaper aims at a target group and relies on having regular readers (i.e. customers), journalists are not as free to express attitudes but have to cater for, or at least consider, their readership's expectations. Evaluative language in this particular discourse is thus not as much an indicator for hierarchical structurers as for seeking closeness or solidarity, and will therefore be treated under this subdimension. Apart from the already mentioned use of titles and forms of address, other operationalisations in this context are the use of adverbs as minimisers or boosters (Eggins and Martin 1997, 232), high frequencies of modals (Fowler 1991, 64) and the inclusion of adjectives and adverbs with strong positive or negative connotations (Biber 2006b, 87-90; Bartsch 2009, 112).

While Martin's distinction between involved and uninvolved settings again mainly focuses on the concrete prior contact of the communicative partners and is therefore not easily applicable to written texts, Poynton (1989) suggests a differentiation between task-oriented and person-oriented, thereby relating to the potential orientation of goal types mentioned by Halliday and Matthiessen (2014, 41; see section 3.2.1.2). Neumann (2013) distinguishes neutral, casual and consultative style to map differences in distance. However, especially in Neumann's study texts from very different registers were analysed which could be expected to show huge differences with regard to social distance; within the closed field of newspaper discourse, diversity can be expected as well, but to a lower degree. Any categorisation of potential styles should therefore be regarded as continuums rather than separate categories with clear boundaries and should not be too complex. The present study will trace differences along a simple scale of casual as opposed to formal style.

For this, casual style will be represented by the above-mentioned instances of evaluative language, namely use of titles and forms of address, modality, minimising and intensifying adverbs and strongly connoted adjectives and adverbs. Concerning the first, minimisers and boosters, lists of potential adverbs with these functions can be found in different studies. Eggins and Martin (1997, 231-32) suggest *really*, *pretty*, *very*, *totally*, *more* and *only*, Matsumoto, Hwang and Sandoval (2015, 232) mention *just, simply* and *merely* as well. Benamara et al. (2007, 2) add stronger expressions like *exceedingly*, *extremely*, *immensely*, *barely*, *slightly* and *scarcely*, and Biber et al. (1999, 564-567) add, among others, the items *nearly, relatively, fairly, entirely* and *fully*. Further examples are *hardly*, *absolutely*, *quite*, *thoroughly*, *utterly*, *completely*, and *most*. Of course, the list is not complete here, and other adverbs can have intensifying or minimising functions depending on the context in which they are used; the above-mentioned ones are quite straightforward however and can therefore be analysed quantitatively without risking the inclusion of too many false hits. For every one of them, the results will of course be checked for negations, which would turn the meaning around. Additionally, *more* and *most* of course often realise comparatives and superlatives, which is why these two forms of adverbs will be included in general.

The second operationalisation, strongly connoted adjectives and adverbs, will rely on maximally unambiguous lexical items as well, however, predefined lists for this purpose would risk being either extremely extensive or too selective. The selection for this investigation will therefore draw on the data itself. The corpus contains 7,389 different lemmata for the part of speech of adjectives, 2,466 of which occur only once. The 716 most frequent lemmata, which represents every item that occurs thirty times or more, make up 76% of all adjectives in the corpus. A similar distribution can be found for adverbs: there are 1,299 different lemmata for lexical adverbs (in contrast to dominantly functional adverbs, see section 4.2.2), of which the top 60, that is every lemma with 150 or more instances, contain 76% of the total number. From these 716 adjectives and 60 adverbs, the most strongly connoted ones will be filtered manually. Although this is a subjective approach, it is a more thorough way to process these amounts of data than drawing on pre-collected lists the appropriateness and completeness of which for the case of newspaper discourse could not be guaranteed, and not tested without manually checking. A manual selection is therefore preferred at this point. In order to keep the subjective influence as low as possible, as was already said only the most clear-cut items will be selected.

In addition to these operationalisations of evaluative language, six other linguistic characteristics which have been found to represent closeness will be analysed: frequent place references as markers of a shared background, contractions, frequent use of 1st and 2nd person pronouns, low use of passive voice, and low frequencies of nouns and nominalisations (Biber 1988, 73-75; Steiner 2004, 18; Bartsch 2009, 112; Neumann 2012, 86; 2013, 67-71). In this investigation, the latter will be represented in terms of gerund-forms with a nominal part of speech and words with the derivational suffixes *-tion*, *-ness*, *-ity*, *-ism*, and *-ment*, always of course including also their plural forms (Biber 1985, 344; Neumann 2013, 128). Casual style can be seen as reflecting a low social distance as well as a tendency to be rather reader-oriented. In contrast to this, formal style as the choice to signal a high social distance will be represented by negative values for the indicators for casual style.

3.2.3 Mode of Discourse

The third and last parameter to define the context of situation in a systemic functional sense is that of the mode of discourse. It represents "what part the language is playing, what it is that the participants are expecting the language to do for them in that situation" (Halliday and Hasan 1985, 12). To capture this more concretely, Halliday and Matthiessen (2014, 33) define six aspects of mode, which partly overlap with the subdimensions of field and tenor as well as with each other, making an empirical approach difficult. Neumann (2013) and Steiner (2004) therefore subsume their contents under three headings, those of *language role*, *channel* and *medium*, which will serve as subdimensions in the present study as well.

3.2.3.1 Language Role

The subdimension of language role defines how important language is in a given situation in order to complete the social action successfully. It is therefore linked to the concept of goal orientation (cf. Hasan 1999, 281-82), however, we can reach goals in different ways and using different techniques. Language is just one option; we can also use gestures or draw pictures or rely on the institutionalisation of a situation (if, for instance, pupils always stand up at the beginning of a lesson when their teacher enters, the mere action of entering will convey the message to the pupils). The goals the participants pursue therefore have an influence on the language role, but do not entirely determine it.

In most studies that look more closely at language role, the dimension is treated as a scale ranging from ancillary to constitutive (Halliday and Hasan 1985; Hasan 1999; Steiner 2004; Neumann 2013). If language plays an ancillary role, it is used as an additional, supportive means to complete the action, while a constitutive language use implies that the action is completed only by verbal interaction. Concerning newspaper language, it is safe to assume that the role language plays tends strongly towards the constitutive end of the scale, as newspaper articles mainly consist of writing. Nevertheless, other features like pictures and, in the case of online editions, videos are often embedded and complement the purely linguistic description of the story. Formatting, too, can be taken into account here, as bold or large print has implications beyond the linguistic level. To adequately analyse these non-verbal aspects of newspapers however would require a multimodal corpus. Although operationalisations have been suggested to analyse the language role mainly on the basis of linguistic characteristics (Neumann 2013, 208-10), in the present study such an examination would only allow conclusions about the potential presence of other means next to language, but give no reliable information about their nature or their content. Therefore, language role will not be included in the analysis.

3.2.3.2 Channel

A similar case is the subdimension of channel. The channel represents "the physical conditions of the communication" (Neumann 2013, 75) as is traditionally defined as either graphic or phonic (Halliday and Hasan 1985, 58; Hasan 1999, 282; Halliday and Matthiessen 2014, 33). Steiner (2004, 19) makes further distinctions between more specific channels such as telephone calls or electronic writing, which is also mentioned by Neumann (2013). Martin (1992, 511; see also Eggins 1994, 54) on the other hand presents two variables along which channels can be defined, namely the degree of visual and that of aural contact between the participants. This allows a more fine-grained determination of channels than distinguishing between phonic and graphic only, yet for newspaper writing, the difference is irrelevant. In terms of Martin's model, both visual and aural contact between reader and writer are zero, and in the more simplified sense the channel is purely graphic. Only the above-mentioned videos which can be embedded in online news are an exception to this, as they add a phonic element. It is disputable whether they change the contact dimensions between the parties involved however; if the journalist is regarded as the sender even while the reader plays an embedded video, the dimensions in Martin's model do not change. If, however, the video contains direct speech from other parties, one might argue that the dimensions change because the role of sender changes to a new person.

In the present study, all articles were collected in the same way, from the newspapers' websites (see section 4.1.2). Variation with regard to channel, even if the more 'modern' distinction between phonic, graphic and electronic is applied, is therefore excluded by the sampling technique. Embedded videos are the only element for which a switch in channel would be possible; since these are not included in the corpus, channel will be left out of the analysis.

3.2.3.3 Medium

The most relevant aspect of mode for this study is that dealing with the medium of the communicative situation. Essentially, the medium can be spoken or written, but is to be found on a more abstract level than the distinction between phonic and graphic channel (Halliday and Hasan 1985, 58; Neumann 2013, 75-76; Halliday and Matthiessen 2014, 33). In more conventional terms, the medium can be said to reflect the style, and crossovers between medium and channel happen frequently; not everything that appears in a graphic channel follows the typical style of written language (examples are dramas, chat room conversations or semi-scripted radio talk), and in turn, texts can be transmitted via a phonic channel but show characteristics of written language (as for instance speeches or radio news). While the dimension of channel is absolute and a text either has phonic or graphic attributes or not, the notions of spoken and written medium work as end points of a scale along which texts can be located.

As was said in the previous section, the texts in the corpus for this study all appear in the graphic channel and significant variation can be excluded; with regard to medium however, this is not the case. Journalistic writing is a very conscious process and the product goes through various editing steps before it is cleared and published. Tendencies towards either end of the medium scale are not random but contribute to the text and reflect the context in which it was produced, which is why differences in medium as well as the range of diversity in this aspect are of interest for comparing both domains and varieties. A lot of studies have analysed characteristics of written and spoken language and have found differences on many levels. A lot of these differences of course originate in the mere structure of the situation, such as turn-taking or interruptions in spoken language. Others are phonological in nature, like the use of intonation, speed and paratones in speech.

In these aspects, the texts in this study cannot vary, however, differences can also be observed on the lexico-grammatical level, on which the texts are comparable.

The most prominent differences can be found with regard to the lexis; spoken language in general contains more pronouns, contractions and sentence-initial conjunctions, while written language tends towards nouns and nominalisations (Eggins 1994, 56-57; Biber et al. 1999, 81-92, 333; Steiner 2004, 20; Neumann 2012, 78; Neumann 2013, 75-78). Especially pronouns are usually regarded as representatives of the broader category of references, which is used very differently in spoken and written language. Halliday and Hasan (1976, 31-37; 1985, 76) distinguish between exophoric and endophoric references, the former referring to aspects of the situation, the latter to aspects within the text. Typically, spoken language can be assumed to contain more exophoric references than written language, especially when the communication is face to face and the participants therefore share the dimensions of space and time. Words like *that*, *there*, *these* or *this*, which are named as typical realisations for exophoric references (Halliday and Hasan 1985, 35), can in such a context be underlined with gestures or facial expressions, making the reference clearer and less ambiguous. In most written discourse the participants do not physically share the situation though, which makes exophoric references more difficult.

There are certain aspects in newspaper writing which can be regarded as exophoric references, although pronouns or determiners are unlikely to fulfil this function. Instead, features like hyperlinks, phone numbers, email- or postal addresses and references to prior articles on the subject are exophoric in nature. However, in contrast to the examples given by Halliday and Hasan (ibid., 75-79), they do not refer to something in the immediate situation, in the shared space of the reader and writer, since there is none beyond the text. Rather, they refer to something which is in neither of the participants' spheres but in a third space which is accessible for both. Because of this property, they cannot be said to reflect spoken medium per se, but are a characteristic of newspaper language that serves the purpose of self-advertising (in the case of references to other articles) or providing further information. This distinction between references will therefore not be taken into account in the analysis.

Apart from this, the lexico-grammatical characteristics mentioned above are reflected in the lexical density, which is generally higher in written than in spoken texts (Ure 1971, 445-46). Neumann (2013, 76), drawing on Halliday (1994, 2001), argues that this value itself is not ideal since it does not take into account language-specific characteristics with regard to the usage of function words. Instead, it is suggested to calculate the density of lexical words relative to the number of clauses. This is indeed a valid criticism, but since the corpus used here is not annotated with clause structures, such a calculation is not possible for the whole set. Since the present study looks at varieties of one language instead of different languages however, this flaw in the measurement of lexical density it is less crucial. Nevertheless, it will be combined here with a more detailed examination of the distribution of word classes in the texts, to see whether high or low values in lexical density do indeed reflect the general relation between function and content words or whether they originate from extreme values of a particular word class which might skew the representation.

3.3 Summary

Based on the argumentation in the previous sections, the present study will include four subdimensions – the experiential domain for the field of discourse, social distance and social role relationship for the tenor of discourse, and medium for the mode of discourse. Appraisal or affect, which is sometimes counted as a separate subdimension of the tenor, will furthermore be included in the analysis of social distance. An overview of all operationalisations is given in Table 3.1:

Table 3.1: Operationalisation - Overview

Parameter	Subdimension	Linguistic indicators
Field of Discourse	Experiential Domain	Vocabulary Tense Expressions of time Lexical density
	Goal Orientation	Type-token-ratio Mood Pronominal use
Tenor of Discourse	Social Role Relationship	Mood Modality Amount of keywords Mean word length
	Social Distance	Evaluative language - Use of titles and forms of address - Boosters and minimisers - Evaluative adjectives and adverbs - Modality 1^{st} and 2^{nd} person pronouns Place references Contractions Voice Nouns and nominalisations

Mode of Discourse	Medium	Pronominal use Sentence-initial conjunctions Contractions Nouns and nominalisations Lexical density

In some cases, the operationalisation has been adapted to the discourse of newspapers, as became apparent especially in the separation of place and time references and the assignment of evaluative language to the subdimension of social distance. Although this decreases the comparative value to other discourses on a lexico-grammatical level, it does take into account that the linguistic realisation of the parameters and their subdimensions differ. On the more abstract level of situational context, a comparison is therefore possible, while the operationalisations reflect discourse-specific particularities and thus ensure an adequate representation of the abstract concepts.

The operationalisation will be the foundation for the analysis and interpretation in the second part of this work. Drawing on the theoretical background that was given in chapter 2, it is to reflect newspaper language and its various contextual characteristics in a systematic, quantitative way. Before the actual analysis and the results for the features are described however, chapter 4 will specify the dataset on which they are based as well as describe its advantages and limitations.

4 Corpus Design

In the previous chapters, the collection for the present study has been mentioned several times already, and section 2.3.3 introduced the newspaper landscapes of the five varieties and the newspapers included in the dataset. It is the aim of this chapter to describe the corpus in more detail and explain the collection technique as well as sampling decisions that had an influence on its compilation.

The first section will start with a general description of the data and what it is meant to represent. It will then go on to define the parts of the corpus – for the regional varieties, national borders provide basic definitions, but the newspaper domains are more abstract and will therefore require more exact descriptions to eliminate ambiguity for the subsequent analysis. The section will then turn to the question of representativeness, which is highly disputed among scholars working with large corpora and will be defined for the collection of this study. The second section of the chapter will shift the focus to the technical aspects of the corpus, including the annotation and mark-up and the processing software. To summarise the information, section 4.3 will provide a compact overview of the whole dataset.

4.1 Data Collection

As was already described in section 3.1, a quantitative approach holds many advantages when analysing language variation comparatively along functional and regional lines. The main advantage certainly is a purely pragmatic one – to compare language across different domains and regional varieties and at the same time ensure their adequate representation in the analysis, a large amount of data is necessary and makes automatic processing the most reasonable choice. Although the decision to use a quantitative approach was therefore easily made, it gave rise to the question of the exact data on which to base the analysis.

Studies into newspaper language are numerous, and every empirical work requires data of some sort. Most qualitative studies rely on small datasets or even individual articles which are not meant to be representative and do not offer themselves as foundations to large-scale quantitative work. However, corpus-based studies of newspaper language can be found as well, and therefore, larger datasets of articles have been compiled before.

Not all of the existing corpora that include newspaper articles are newspaper specific. The ICE for instance, which has been mentioned a number of times before, contains newspaper texts, but only distinguishes between reports and editorials. Similar restrictions hold true for the British National Corpus (2007) and the Corpus of Contemporary American English (Davies 2008), which furthermore are specific for one regional variety only. This is also the case for most collections that are more specialised on newspaper language; the American National Corpus (2015) contains a subcorpus for US-American news language, and Macalister (2001) compiled a corpus of newspaper articles from New Zealand. Furthermore, there are two compilations built for diachronic analyses of British newspaper texts, the Zürich English Newspaper Corpus (Fries and Schneider 2000) and the Rostock Newspaper Corpus (Schneider 2012). The biggest contemporary news-specific corpora are the Reuters Corpus with roughly 810,000 Reuters news items (Reuters 2000) and sub-parts of LexisNexis (2016), which contains English-language newspaper articles from all over the world. Although LexisNexis can also distinguish articles based on their topic, it does not include any part of speech-tagging, which makes queries on the lexico-grammatical level very difficult.

Although some of these collections would have been accessible, none of them would have served the purpose of a comparative analysis across domains and varieties. Apart from this, the documentation of sources was not always sufficient, and information on the publication date, newspaper or author was often unavailable; yet in order to represent the newspaper discourses of the different regions properly, a variation in newspapers as well as authorship should be guaranteed. For these reasons, a new corpus was compiled which is tailored to meet the requirements of the research questions of this study.

The decision to compile a corpus triggered issues concerning its exact target population and design as well as the implications of representativeness. To ensure transparency of the analysis and interpretation in the following chapters, the next section will define the target population and sampling frame and thereby provide an insight into the collection and its proportions and content.

4.1.1 Target Population

The target population of a data collection represents the entity which is of interest in a study. While in sociological research this is usually a certain group of people, in corpus linguistics it specifies the language that is to be analysed (Biber 1993a, 243; Springer 2010, 100). Biber (1993a, 243) criticises that many corpus-based studies do not consider their target population sufficiently and therefore sample data that is not adequate for the respective research goals, which in turn leads to false statements and overgeneralisation of the results. In order to prevent this, he suggests defining two important aspects of the target population: "(1) the boundaries of the population – what texts are included and excluded from the population; (2) hierarchical organisation within the population – what text categories are included in the population, and what are their definitions". These two aspects are indeed very helpful for the description of the target group, as they capture the relation between the texts both on a vertical and a horizontal level. This way they ensure that the population is defined clearly enough to be operationalised for the sampling frame in the next step, but they also mirror potential subcategories that can be important in a study.

Roughly speaking, the corpus for the present analysis is meant to represent present-day newspaper language from different varieties of English as well as different newspaper domains. It was furthermore defined that the varieties should encompass both native and L2 varieties, since especially in the latter the developmental status is less clear and often under dispute due to the differences in language policy and status. Similarities or discrepancies in comparison to different native Englishes can help determine a variety's development and linguistic particularities more easily, and covering both in the corpus thus allows the represented varieties to function as control groups and reference corpora for each other as well as being an element of research themselves.

Concerning the parameter of newspaper language, it was defined that thematic diversity should be taken into account instead of treating all newspaper texts as one group. This is necessary in order to analyse functional variation which is regarded in this study as a measurement of the developmental status. Comparing newspaper language in general across the varieties would have shown differences, but it would have been impossible to say whether some newspaper domains show more linguistic particularities than others, or whether there are universals in any of the domains. Furthermore, the degree of variation within each variety could only have been measured fairly crudely, since there would have been no guarantee that the components for the different varieties contained comparable distributions of domains. Low or high internal variation could have been the result of an overrepresentation of a certain domain, which is avoided by including this dimension of variation. This reflects Biber's second aspect of defining the target population of a study, since the newspaper domains are subcategories of newspaper discourse. The same holds true for different newspapers and authors, which were to be taken into account to avoid overrepresentation of individual styles and media traditions.

These boundaries and definitions of the target population in the next step determined the sampling frame and eventually the data sampling. While the target population is still relatively broad, it was put into action more concretely in these phases, which will be described in the next section.

4.1.2 Data Sampling

The purpose of defining a sampling frame is to operationalise the specifications of the target population, and thereby to determine how these abstract specifications can be represented properly in a dataset. For the target population described above, this meant deciding on adequate regional varieties and newspaper domains, concretising the implications of "present-day" and setting clear boundaries to avoid the abovementioned overrepresentation of individual styles.

As determined for the target population, the dataset was supposed to include regional varieties with different sociolinguistic histories and political status in the form of both native and L2 varieties. To allow cross-comparisons and at the same time create control groups, it was determined that at least two of each kind should be sampled. Especially with regard to the L2 varieties some restrictions were given by the accessibility of newspaper data or the presence of English newspapers in general. Of those regions where sampling appeared possible, Kenya and Hong Kong were picked for several reasons. As was seen in section 2.1.3, the Englishes in both regions share their colonial background, yet the language was originally introduced very differently; in Kenya, the spread of English was not intended and only few people were taught by the colonial settlers while in Hong Kong, English was introduced more radically and widely. Furthermore, Hong Kong was connected to British rule for much longer and was eventually handed over to China peacefully, in contrast to Kenya, whose strive for independence was much more troublesome.

This resulted in differences concerning language policy. In Hong Kong, English-speaking universities and schools were founded and due to the region's economic importance, being proficient in English was desirable for many people for merely professional reasons. Although to a certain degree this holds true for Kenya as well, Kenya is much larger than Hong Kong and displays more internal diversity in terms of the geographical distribution of population and economic density and socioeconomic factors; Nairobi, the country's capital, is the economic centre, but in large, especially rural areas of the country the local languages are dominant and English is less relevant except for official purposes.

In both regions, the British influence has certainly been significant, but it was enforced differently. Although in terms of Schneider's (e.g. 2003, 2007, 2014) model, both varieties display features of nativisation, being located in phase three, and the status of official language is given in both cases, it seems too rough to map them onto the same circle in for instance Kachru's (e.g. 1985, 5) model. Kenyan and Hong Kong English were therefore determined as examples of L2 varieties as a closer comparison of the two, especially in relation to native varieties, seemed promising to fulfil the requirements set by the definition of the target population.

With regard to the native varieties, the decision which ones to include was a lot easier. For one, there are much fewer native than L2 varieties, and also fewer limitations due to data accessibility – with few exceptions, all newspapers are published in English and most of them feature online versions as well. Furthermore, British English was set as it constitutes the origin of English in Kenya and Hong Kong and was the main linguistic influence during their respective colonial times. US-American English was included due its influence today; large parts of current popular culture, including television series, music, movies and literature, is produced in the USA, which makes an influence of US-American English on other speakers of English probable. In addition, the country's political importance and involvements make it an issue of interest in news everywhere; since the USA is therefore written about in the other countries included here, it seemed reasonable to include it as well to trace the influence and put the US-related news from the other varieties into perspective.

As a last variety, Australian English was included. Since the developmental status of the varieties from Kenya and Hong Kong is one major aspect of this study, a comparison to the dominant standards of British English and US-American English did not seem sufficient; an L2 variety can display more similarities with one than with the other, but it can also appear distant from the two. This does not necessarily mean that there is no native influence however. Especially in the case of Hong Kong, Australia is much closer in geographical terms than Britain or the USA, and for Kenya, too, Australia is closer than the United States, which strengthens the potential news worthiness in terms of proximity. This is relevant for Britain and Australia as well, since the countries are politically connected via the Commonwealth of Nations. Australian English therefore constitutes a potential influence or reference point for the varieties in question, and was included as well.

Of course, all of these varieties feature further internal regional varieties in the form of dialects and accents. Since an analysis of these internal diversities would have been beyond the scope of this work however, they are treated as national entities. To ensure that the data does not show too drastic linguistic particularities of individual regions, only newspapers that are published nationwide were considered for this study.

Section 2.3.3 already provided a detailed overview of the newspaper landscapes in the five regions and listed the largest papers as well as the sampled ones respectively, which is why this will not be repeated at this point. Independent of the regional variety, the newspapers for this study were selected on the basis of three criteria: they had to be published in English, they had to be freely accessible at the time of the data collection, and, as was said above, they had to be published nationwide. Furthermore, the collection was designed to reflect potential differences between popular and quality press by balancing the two newspaper types. As a definition of these categories is very difficult and hardly comparable across the regional varieties however, this differentiation has to be regarded with caution and cannot be used as a basis for conclusions (Fest, forthcoming). For British and US-American newspapers, common assignments of the most prominent papers can be found in previous studies, but especially in Kenya and Hong Kong, where few English-language papers were available in the first place, separating them in this regard would be very questionable. Since the newspaper titles are available as meta-information in the data (see section 4.2.1), similarities and differences will be traced between individual papers rather than between categories, and the labels 'popular press' and 'quality press' will be used sparingly and with restriction to well-researched cases.

The last major aspect defined in the target population was the inclusion of different thematic news domains. As all papers were to be represented equally (see section 4.1.3 on representativeness), only those domains were included that were present in all papers, which resulted in the categories of economy, hard news, politics, sports, and lifestyle. Not all these labels are self-explanatory, and there are various definitions that allow room for interpretation.

The most clearly defined of these five domains is that of sports news. Not only is the topic very specific, but in printed editions it is also often separated from the other domains by being allocated its own pages in the paper. Apart from the news items themselves, this part can feature tables, statistics, match plans and other non-textual elements such as videos or game animations. The articles usually deal with one particular sport, which makes this domain rather diverse considering the many sorts of sport that can be covered, but also fairly clearly structured. Furthermore, many sports are alike in basic elements and thus result in basic journalistic patterns: news items dealing with team sports for instance, such as football, soccer or basketball, often contain line-ups, transfers and descriptions of actions of various players, while in sports such as tennis or golf, the focus shifts to individual players or tournaments (cf. Fest, forthcoming).

Despite these internal variations and the rather specialised studies that have been done into particular sports, sports news is considered one thematic domain in this study. Also, the collection of articles in this category did not imply any stratification on the basis of the individual sports, which means that the subtopics in this domain are not equally represented. Such a sampling would have been made impossible due to the differences in popularity individual sports hold in the different regions and the respective numbers of articles dealing with them; on the other hand, a sampling reflecting the popularity alone would have required a broad and longitudinal study of the representation of sports in newspapers and the weighing of topics, which would have been beyond the scope of this work.

Much more diverse in its content and yet nearly as distinguishable as sports news is the section usually labelled "Lifestyle" or "Life & Style". The range of topics that are covered here is very extensive and reaches from health questions, social issues and personal stories to restaurant tips, travel advice and fashion reports, to name but a few (Harcup 2005, 119). Despite this diversity, a pattern is clearly recognisable looking at the news values that are emphasised in feature articles. Reports and stories that are published here are independent of, or at least not strictly bound to, particular events or points in time, which makes the value of recency, which is crucial in news in every other section, much less important. Many articles, especially those dealing with "beauty tips" or health advice, might not be out of date for months after their publication. Instead of the aspect of recency, a personal involvement with the reader is often actively sought and accentuated to ensure the relevance of this news section.

The other three sections that make up the corpus, namely economy, politics and hard news, are not always as easily distinguishable as the two prior ones. The most clear-cut category here certainly is represented by hard news, also sometimes referred to as spot news, although an exact definition is hard to find (cf. Reinemann et al. 2012, 223-24). Some scholars rely on individual criteria to distinguish hard from soft news such as timeliness (Shoemaker and Cohen 2006) or angle (Curran et al. 2009), and yet most use the topic for this categorisation and many subsume stories regarding economy and politics under this heading as well (e.g. Schneider 1999; Ungerer 1999; Granato 2002; Curran et al. 2009), yet these two labels could be found separately in every paper and were therefore kept apart. For the present study, the domain of hard news was thus defined rather restrictedly and consists of articles dealing with accidents, natural disasters, crimes and trials, as well as topics such as strikes, protests or reports on social issues. Especially articles of the last category can sometimes come very close to what is being treated in features, and crimes or trials can be linked to any other section, depending on who is involved. For this reason, hard news is rarely found in a separate section of a newspaper, but are most often just labelled "news". Also, since recency is of utmost importance in these articles, events are being reported as quickly as possible, even if some information, such as the circumstances of an accident or the exact number of victims in a crime, is still missing. This results not only in articles being shorter (see overview in section 4.3), but also in the frequent production of follow-up articles depending on whether and when more details concerning the event become known.

Recency is also a relevant news value for economic and political news. In both cases however, the popularity of the parties involved strongly plays into the news worthiness of a story or an event. Big names interlace with the significance of the action itself, an aspect that these news stories share with sports or celebrity news. In contrast to these though, the topics treated in news dealing with economy or politics are largely of a more serious, if not sinister nature. Elections of course are a common subject for political reporting, alongside parliamentary issues, summits and other meetings and debates. News items treating economy frequently contain reports about stock exchanges, shareholdings, company closures or fusions as well as statistical evaluations of and news on the labour market.

It is important to bear in mind that in terms of content, these domains are not always clearly distinguishable, since many topics fall into two or more categories. Especially economic and political news often overlap and also sometimes fall into the range of hard news. Sports news, too, although more isolated, can contain elements of economic or criminal nature, and the area of lifestyle can deal with topics of other domains from a particular perspective, such as offering advice on how to save money or describing politicians' personal life. To avoid any subjectivity with regard to the content and thereby biasing the analysis, the articles were thus not sorted into the thematic categories by the collectors, but were taken as they had been flagged by the newspapers.

Representing the last element of the target population, the sampling frame for "present-day" newspaper language was determined as including articles from the year 2000 or later. A diachronic perspective is not the focus of this work though, which is why the date of publication was not further defined and years are not represented in equal parts. Especially for Kenya and Hong Kong, archives were often not accessible for more than a year back and the parts contain a much higher number of younger articles. In general, all articles were published between 2000 and 2013, with the more recent years being clearly dominant.

In order to avoid the dominance of any one newspaper, the collection aimed at including as many papers as possible. The limit of four papers per variety was defined by the part of Kenya, for which this was the maximum number of newspapers available in English. All articles were collected from the online editions of the papers, since not all newspapers were readily available in their print editions and the publication format was meant to be constant for all articles.

Furthermore, the sampling frame was restricted to containing no more than three texts per author, a boundary which aimed at preventing heavy influences from individual journalists. This limit was applied to the whole collection with the exception of the sports domain for the Hong Kong newspaper *The Standard* – while ten articles could be collected which gave the names of their respective authors, the other 30 articles in this part were attributed to the news agencies Reuters (14 articles) or

AFP (16 articles). As no other articles were available, they were included in the sample.

Within the frame boundaries, the articles were sampled randomly.[18] In total, the final corpus contains 4,000 articles which are evenly spread over the domains, regions and newspapers. For every region, every newspaper contributed 40 articles per domain. This number was not set prior to the collection process but was the limit that could be sampled within the frame boundaries defined above:

Table 4.1: Corpus design - Number of articles per category[19]

	AUS	HK	KEY	UK	USA	∑
Economy	160	160	160	160	160	800
Hard News	160	160	160	160	160	800
Lifestyle	160	160	160	160	160	800
Politics	160	160	160	160	160	800
Sports	160	160	160	160	160	800
∑	800	800	800	800	800	4,000

This way of stratifying the collection is one way to ensure representativeness of the defined target population, but opinions differ as to what makes a dataset representative. Before the processing of the data is explained in more detail in section 4.2, the following section will therefore briefly discuss the different approaches and summarise the decisions with regard to representativeness that were taken for this study.

4.1.3 Representativeness

The steps described in the preceding sections, the definition of the target population and the derivation of the sampling frame, are preferable to a random collecting of data because they help to ensure that the final corpus is representative of the language it is used to analyse. To Leech (1991, 11), this is the quality that sets a corpus apart: "the difference between an archive and a corpus must be that the latter is designed for a particular 'representative' function". Even with clear-cut definitions of the two however, there are open questions regarding the amount and distribution of the data in the corpus.

---

[18] In order to keep the size of the table manageable, not all individual articles included in the sample are listed. The exact composition of the corpus is available upon request.
[19] When required for the sake of readability, the varieties will be abbreviated as AUS, HK, KEY, UK and USA and the domains as ECO, HN, LIFE, POL and SPO in the figures and graphs throughout the analysis. The individual 25 cells resulting from the five varieties and five domains will be referred to as *categories*.

The first issue to consider is that of corpus size. There are different suggestions regarding the ideal amount of data for a corpus; according to Biber (1990, 261; 1993a, 251-52), samples of 1,000 words of their respective text category produce similar results, which is why he considers a sample of 2,000 words or more sufficient to guarantee an even distribution of linguistic features and therefore be regarded as representative. For newspaper language, this text length is above average however, and selecting only articles with 2,000 words or more would rather lower the representativeness of this discourse than raise it. On the other hand, including articles independent of their length would mean that a sample of this size would have to contain more than one text, which would make the corpus parts less comparable as the number would vary.

In contrast to Biber, Oostdijk (1988, 20) assumes that a sample should contain at least 20,000 words, especially if the corpus is to be used to study varieties. In addition, the collection "should comprise a variety of samples which differ with respect to a number of extralinguistic variables", in this case region, news domain and newspaper. Although Oostdijk does not give any definite answer to the question of how large a corpus in total should be, she is positive that small scale collections, if based on a well-defined subset of a language, are suitable and sufficient for a proper analysis (ibid., 14). This view is partly supported by Atkins, Clear and Ostler (1992), who do not discuss the required corpus size as much as the logical relation between target population and sampling frame. In this context they come to the conclusion that the more specialised the target population and language in question, the easier the sampling, as fewer texts meet the requirements in the first place. In general, they take a more pessimistic attitude towards representativeness, stating that any corpus will always be biased and complete representativeness is not possible (ibid., 5).

Despite this concern, the authors discuss not only sampling methods but also, in contrast to Biber and Oostdijk, explicitly say that words are not the only measurement by which to normalise a collection, but that sampling is also possible on the basis of for instance sentences and texts. This is relevant for the study at hand, since articles vary greatly in terms of length and patterns might emerge that are typical of domains or varieties. Normalising by words would have meant losing this information; instead, the corpus used here was normalised on the basis of articles, and all articles were taken over completely, not in the form of excerpts. As was said in section 4.1.2, as many texts were collected as was possible within the frame boundaries.

This sampling resulted in 4,000 articles in total, a collection in which all extralinguistic variables – domains, varieties, newspapers – are represented by equal amounts of texts (see Table 4.1). This way of stratifying the corpus does not take real-life proportions into account; in most newspapers, the domains included here do not occur in equal measure, and the varieties have very different numbers of speakers and therefore also different potential audiences for the papers, which in turn have different circulation numbers. In a proportional sample, these differences would have to be taken into account and reflected in the amounts of articles included. Biber (1993a, 247) points out that although this sampling technique is useful for demographic studies, for corpus compilation it is almost impossible to put into practice. An analysis of these real-life proportions would have to precede the linguistic analysis to define the texts to be sampled, and would then most likely result in some parts to be represented to such a small extent that they could not be analysed adequately in quantitative terms:

> A simple demographically based sample of language use would be proportional by definition-the resulting corpus would contain the registers that people typically use in the actual proportions in which they are used. A corpus with this design might contain roughly 90% conversation and 3% letters and notes, with the remaining 7% divided among registers such as press reportage, popular magazines, academic prose, fiction, lectures, news broadcasts, and unpublished writing. (Very few people ever produce published written texts, or unpublished written and spoken texts for a large audience.) Such a corpus would permit summary descriptive statistics for the entire language represented by the corpus. These kinds of generalizations, however, are typically not of interest for linguistic research. (ibid.)

For a corpus focusing on newspaper discourse, such an extreme variation in terms of distribution is not likely, yet the actual proportions of all extralinguistic variables are nearly impossible to determine and a respective corpus design would not allow comparisons between the language use in the individual parts. Based on proportional sampling, the largest part by far would be the US-American one as the country with the largest population, while the Hong Kong part would be about 45 times smaller. In Kenya, the *Daily Nation* would have to contribute 48% of the articles based on its relatively high circulation numbers, while the two smallest papers together would contribute less than 22% (see also Hunston 2002, 28-30). Proportional stratification can thus be said to represent the actual use of regional and functional varieties, for an analysis and comparison of these factors however it is counterproductive and was not applied for the corpus used here.

To sum up, this investigation is based on a corpus that has been sampled on the basis of language-external criteria, and reflects all its parts in a non-proportional, evenly distributed way. The unit with regard to which the parts are normalised is the text, not the word, which means that the parts vary in the amount of tokens they contain (see Table 4.2). The samples are as large as the frame boundaries permitted, and within these boundaries, the articles were sampled randomly. Since the research aim of this paper is a comparative analysis of both functional and regional variation, this design was chosen as the most representative of the target population and sampling frame.

After the completion of the sampling, the texts were processed further and enriched with annotation and meta-information to facilitate the analyses on the lexico-grammatical level. Since the results presented in the next chapter will rely on these additional features, the following sections will provide details about the text processing and discuss technical issues that occurred in this work phase.

4.2 Data Processing

The data for this study was processed in several steps. All articles were collected from the respective websites and given a text ID in the format [Domain]_[Region]_[001-160]. The collection of each text comprised headlines and the actual article, but no other elements such as comments, pictures, videos, graphs or their respective captions. The individual texts were then enriched with meta-information on their publication as well as linguistic annotation, which will be explained in the following sections.

4.2.1 Non-linguistic Annotation

The non-linguistic annotation that was added consists of a header which was integrated into each text and several attributes concerning the textual structure. The header was included automatically using a tool which relied on the meta-information documented during the collection and saved in a spreadsheet and transferred them into an XML format (Hansen-Ampah 2013):

```
<article title="Tech company eyes bigger slice of emerging markets"
author="Charles Wokabi" publisher="DailyNat" pubPlace="Kenya"
topic="Economy" journalType="Broadsheet" year="2013" text_id="Econ-
omy_Key_012"></article>
```

The header in each article contains its headline (*title*), author, the newspaper from which it was collected (*publisher*), the region it came from (*pubPlace*), the domain from which it was sampled (*topic*), the newspaper type (*journalType*), the year of its publication and the text ID it has in the collection. These attributes stored in the header allow the corpus queries to be matched to their respective values and thus display the results in a more systematic way. The opening and closing parts of the header, <article> and </article>, furthermore function as markers of the beginning and end of the text.

Concerning the textual structure, the corpus contains tags for paragraphs, <p></p>, and mark-up to define headlines. While the former relies entirely on the structure the article had in its original form, the latter raised issues of comparability between the different papers. Every article that was sampled had a clearly distinguishable headline, which was set apart by formatting and sometimes also by colour. This headline is represented in the header as "title", and was annotated in the text as <h1>. However, many articles displayed a second and sometimes even third layer of text that was highlighted in some way, but was often less identifiable than the primary headline. In some cases, the form or positioning of these features indicate its function: if the element in question is set under the main headline but still separated from the rest of the article by a graphical feature like a line, or if it is smaller in font size than the headline, but bigger than the text, it can be considered a subheadline. If, on the other hand, the element consists of a whole paragraph and has the same font size as the text, but is set apart in bold writing, then its function is rather that of an intro.

However, most cases are somewhere in the middle – some articles contain whole paragraphs set in bigger font than the text and positioned right under the main headline, still others feature bullet points in bold or bigger font. Patterns are recognisable for individual papers, which helps common readers who know the structure of 'their' paper to navigate faster (Burger 2005, 114). For a comparative analysis however, a systematic categorisation of these elements was necessary at least in rough terms. The sorting was therefore based on characteristics intros and headlines have been found to display, and on the degree to which the elements in the corpus resembled one or the other.

Subheadlines, as a type of headlines, typically have a syntactic structure that is very different from the rest of the article. It is often elliptic, and the verbs are kept in present tense. It can contain puns and if it includes quotation marks, they are singles instead of doubles (Burger 2005, 115-20; Conboy 2007, 13-16; Bednarek and Caple 2012b, 101). In contrast to this, intros, or leads, contain full sentences and are written in the same tense as the main article. They answer the first questions which are relevant with regard to the story and give more elaborate information on the issue than a headline. In many newspapers that work with intros, it starts with a name or place in capital letters (Burger 2005, 121-23; Bednarek and Caple 2012b, 96-100). Very importantly for the distinction in this corpus, a headline can be followed by an intro; an intro however is a singular element that will not be followed by another one.

Based on these characteristics, the occurrences in the collection were defined as either a subheadline or an intro, and the former were marked as <h2>. Those that were considered an intro were assigned the tag <bold>, however, only those elements that were set apart in formatting were analysed for this distinction. It is very likely that other articles contain paragraphs that meet the specifications of an intro but are simply not set apart visually. While headlines, including subheadlines, will be referred to in some of the analyses, this introductory element will therefore not be treated individually in this investigation.

4.2.2 Linguistic Annotation

In addition to the meta-information and structural attributes that were described above, the data was enriched with linguistic annotation on word level. For this purpose, the annotation software TreeTagger (Schmid 1994, 1995) was used which provides part of speech-tagging as well as lemmatisation. As it includes a separate tag for sentence-final punctuation (exclamation mark, question mark and period) it can also be used for basic syntax-related queries such as words or parts of speech in sentence-initial or -final position.

By default, the TreeTagger works with 43 different tags of which 36 are word-related and seven refer to different items of punctuation or brackets (Santorini 1991; see appendix 1 for a complete list of tags). During the annotation process, a number of technical issues occurred which made alterations necessary; first of all, the tagger failed to assign lemmata and sometimes also part of speech-tags to several more recently emerged lexical items such as *Twitter*, *Facebook* or *blog*. The lexical item *e-mail*, too, caused problems due to different ways of spelling. These errors, since they were systematic throughout the whole corpus, were corrected in an automated way.

A second issue was the diverse use of tokens like quotation marks and hyphens. In most of the articles included here the authors made use of double quotation marks in the text and of single quotation marks in headlines and quotes in quotes. There were very few exceptions to this, and these were standardised to this common format for this investigation in order to ensure that queries would produce results from all newspapers alike. Furthermore, types wrongly used as quotation marks, like the accent marks, were systematically changed to the standard signs as they would have increased the error rate of the TreeTagger.

Next to these changes regarding punctuation and lemmatisation, the above-mentioned tagset of the TreeTagger was enhanced by three tags to eventually include 46. This addition was developed in order to distinguish between adverbs of a more functional nature and those belonging to the group of content words, a distinction which is relevant for the measurement of lexical density. By default, the TreeTagger assigns the part of speech-tags RB (adverbs), RBR (comparative adverbs) and RBS (superlative adverbs), yet adverbs can vary greatly with regard to their actual function. Biber et al. (1999, 552-69) differentiate between eight kinds of adverbs, namely of place, time, manner, degree, stance, linking, addition / restriction and "other meanings". This is a very thorough categorisation which emphasises the complexity of this word class; however, embedding it into a corpus would require a lot of manual input into the annotation and would have been beyond what was possible in this work. Instead, all adverbs (grouped by their respective lemmata) were subjected to the annotation software used on the SUBTLEX-US corpus (Brysbaert, New, and Keuleers 2012), which assigns not one part of speech, but all parts of speech this word has been found to take in the language. Out of these, the dominant one is defined and its dominance expressed in percentages out of all usual uses.

For the distinction to be made here, all those lemmata that showed less than 90% dominance of a content word class were categorised as function words. Of those remaining, 146 had not been assigned anything by the annotator, which was partly due to variations in spelling and hyphenation, but was also caused by the general rarity of the respective words in cases like *sonorously* or *viscerally*. These 146 instances were therefore sorted manually into content and function words.

In order not to mix up the two annotators for the dataset, the annotation by the SUBTLEX-tagger was not implemented as a whole; instead, the tags of the adverbs now categorised as function words were changed from RB to RBF, RBR to RBFR and

RBS to RBFS respectively. This way, the original annotation is still visible and the superlative and comparative forms are kept. Furthermore, despite the diversity within the word class, a query for all adverbs is still possible.

All annotation, linguistic and non-linguistic, is formatted in accordance to the processing software that is used in the analysis. For basic queries and keyword lists, this work will rely on AntConc (Anthony 2005), while more sophisticated queries on the basis of part of speech-tagging and lemmatisation are done using CQP (Evert and the OCWB Development Team 2010). Of course, in each case all analyses of the same kind are done with the same software for all domains and varieties, so that comparability is always ensured.

4.3 Overview

Table 4.1 already showed the general stratification of the corpus and its normalisation on the basis of articles instead of words or sentences. This leads to variation in terms of word number between the parts, which will be taken into account in the analysis by working with relative frequencies whenever quantitative comparisons are being drawn. To provide an overview of the exact distribution of the parts, Table 4.2 lists the number of word tokens (first line) and all tokens (second line) – including punctuation – for the individual domains, varieties, and newspapers.

Table 4.2: Corpus size - Overview

Region	Newspaper	Economy	Hard News	Lifestyle	Politics	Sports	∑
Australia	Herald Sun	13,502 14,970	11,650 13,130	23,538 26,786	18,562 20,741	16,379 18,741	83,631 94,044
	The Courier Mail	18,723 20,806	17,262 19,319	29,065 32,878	17,752 19,779	21,209 23,624	104,011 116,406
	The Sydney Morning Herald	26,418 29,494	15,962 17,886	26,303 30,041	21,295 23,656	24,548 27,800	114,526 128,877
	The West Australian	12,423 13,705	9,908 10,882	22,450 25,412	19,003 21,194	18,034 19,927	81,818 91,120
	∑ Word tokens ∑ All tokens	71,066 78,975	54,782 61,217	101,356 115,117	76,612 85,370	80,170 89,768	383,986 430,447
Hong Kong	China Daily HK Edition	19,212 21,515	8,705 9,780	24,890 28,332	21,179 23,889	19,605 22,173	93,591 105,689
	The Standard	12,024 13,493	14,580 16,226	21,442 24,276	13,131 14,597	9,996 11,223	71,173 79,815
	South China Morning Post	20,615 22,922	13,842 15,466	34,374 39,230	20,365 22,783	22,273 25,015	111,479 125,416
	The Wall Street Journal Asia	30,103 33,728	27,825 31,111	36,896 42,661	27,625 30,521	34,490 38,826	156,939 176,847
	∑ Word tokens ∑ All tokens	81,964 91,658	64,952 72,583	117,602 134,499	82,300 91,790	86,364 97,237	433,182 487,767
Kenya	The Star	13,383 14,676	9,571 10,441	32,830 36,416	14,639 16,011	12,982 14,455	83,405 91,999
	The People	13,592 14,901	16,402 17,935	31,607 35,592	16,978 18,682	12,519 13,669	91,098 100,779
	Daily Nation	19,354 21,308	16,591 18,239	34,229 38,690	23,130 25,525	22,372 25,157	115,676 128,919
	The Standard	19,369 21,506	15,790 17,488	27,750 31,230	20,236 22,317	16,863 18,936	100,008 111,477
	∑ Word tokens ∑ All tokens	65,698 72,391	58,354 64,103	126,416 141,928	74,983 82,535	64,736 72,217	390,187 433,174

UK	The Sun	15,443 17,315	12,926 14,675	26,529 30,154	16,325 18,446	17,552 19,780	88,775 100,370
	Daily Mail	21,866 24,300	23,550 26,695	23,501 26,614	25,156 28,020	24,632 27,642	118,705 133,271
	The Guardian	24,368 27,200	21,344 24,039	36,644 42,006	23,564 26,466	29,436 33,280	135,356 152,991
	The Daily Telegraph	22,869 25,807	24,764 28,173	27,459 31,345	23,500 26,475	28,274 31,596	126,886 143,396
	∑ Word tokens ∑ All tokens	84,546 94,622	82,584 93,582	114,133 130,119	88,545 99,407	99,894 112,298	469,702 530,028
USA	Daily News	16,581 18,904	14,562 16,773	19,999 23,400	14,366 16,429	24,424 27,977	89,932 103,483
	New York Post	17,105 19,345	10,844 12,519	33,850 40,281	11,333 12,998	22,088 25,379	95,220 110,522
	The Wall Street Journal	31,660 35,463	29,437 33,251	39,857 45,700	35,753 40,010	27,504 31,254	164,211 185,678
	USA Today	19,892 22,428	30,454 34,886	23,924 27,545	27,620 31,633	23,687 27,187	125,577 143,679
	∑ Word tokens ∑ All tokens	85,238 96,140	85,297 97,429	117,630 136,926	89,072 101,070	97,703 111,797	474,940 543,362
∑ Word tokens **∑ All tokens**		**388,512** **433,786**	**345,969** **388,914**	**577,137** **658,588**	**411,512** **460,172**	**428,867** **483,317**	**2,151,997** **2,424,777**

5 Analysis

On the foundation of the theoretical background and operationalisation, and applying the methodology outlined in chapter 4, it is this chapter's aim to present the results of the individual analyses. In its structure it will follow the three register parameters of field, tenor, and mode of discourse and their respective subdimensions. The results for the linguistic features defined in chapter 3 will be described separately and discussed for themselves, with a brief summary at the end of each parameter's section. A more encompassing discussion of the implications which takes the results of all variables into account and puts them into the context of variety studies and media language will follow in chapter 6.

5.1 Field of Discourse

The first parameter to capture a situational context, the field of discourse or "nature of the social action" (Halliday and Hasan 1985, 12), will be represented in this investigation by the subdimensions of the experiential domain or "subject matter" (Neumann 2013, 49) and the goal orientation of the discourse. Together the two dimensions include the analysis of seven linguistic features, which will be described in the following sections.

5.1.1 Experiential Domain

The aspect of the experiential domain reflects the topic of a text and has therefore been operationalised in the form of vocabulary, expressions of time, tense, and lexical density. These four linguistic markers require different approaches to their respective analyses, which will be described in detail in the individual subsections.

Vocabulary

The vocabulary used in a text is the most direct indicator of its content, however, it cannot be compared in purely quantitative terms but has to be determined via the most frequent lexical items as well as respective keywords (Fest, forthcoming). As the generation and evaluation of such lists for each of the 4,000 articles would have gone beyond what is possible for this study, these lists were done per domain and variety. Table 5.1 shows the ten most frequent content words per domain counted on the basis of their respective lemmata – a complete list for every domain within every variety can be found in appendix 2.

Table 5.1: Ten most frequent lexical items per domain[20]

Economy							
No	Lemma	Freq.	%[21]	No	Lemma	Freq.	%
1	say	3,808	0.98	6	bank	1,226	0.32
2	year	2,014	0.52	7	new	1,029	0.26
3	market	1,557	0.40	8	billion	944	0.24
4	cent	1,406	0.36	9	million	909	0.23
5	company	1,332	0.34	10	price	891	0.23
Hard News							
No	Lemma	Freq.	%	No	Lemma	Freq.	%
1	say	4,592	1.33	6	court	816	0,24
2	police	2,136	0.62	7	people	782	0.23
3	do	940	0.27	8	take	714	0.21
4	year	933	0.27	9	officer	652	0.19
5	man	900	0.26	10	find	642	0.19
Lifestyle							
No	Lemma	Freq.	%	No	Lemma	Freq.	%
1	say	4,493	0.78	6	people	1,384	0.24
2	do	2,766	0.48	7	time	1,231	0.21
3	year	1,504	0.26	8	go	1,173	0.20
4	make	1,495	0.26	9	also	1,168	0.20
5	get	1,433	0.25	10	take	1,115	0.19
Politics							
No	Lemma	Freq.	%	No	Lemma	Freq.	%
1	say	4,928	1.20	6	minister	1,128	0.27
2	government	1,554	0.38	7	election	1,106	0.27
3	Mr	1,476	0.36	8	year	1,057	0.26
4	do	1,283	0.31	9	also	898	0.22
5	party	1,250	0.30	10	president	879	0.21
Sports							
No	Lemma	Freq.	%	No	Lemma	Freq.	%
1	say	3,263	0.76	6	play	1,371	0.32
2	do	1,842	0.43	7	year	1,371	0.32
3	team	1,627	0.38	8	go	1,225	0.29
4	win	1,504	0.35	9	get	1,186	0.28
5	game	1,378	0.32	10	player	1,179	0.27

This list shows important differences, but also similarities between the five domains of newspaper writing. What is striking is the dominance of the verb *say* in all domains,

[20] In the case of lexical verbs, the part of speech-tagger does not reliably distinguish between different functions, i.e. as a modal or main verb. For the lemma *do*, the tag MD (modal) was assigned only five times, and *have* and *be* are never tagged as modal verbs, which is why the inclusion or exclusion of these items from the list is disputable. Spot tests were conducted which showed that *have* and *be* occur most frequently as modals, whereas for *do*, these cases appeared the minority. In all domains, *have* and *be* were the most common verbs, but were omitted in the lists because of this. *Do*, however, was included.

[21] With few exceptions, figures will be rounded to two decimal places throughout the analysis to increase readability. All calculations were done on the basis of the exact figures.

which hints at the reportative nature which all texts apparently have in common to a certain degree. Furthermore, it is the most neutral verb to accompany a direct or indirect quotation, which makes it preferable to other verbs such as *declare*, *state*, *explain*, or even *shout* or *grump*. Another similarity between the domains is the relatively frequent use of the noun *year*; a closer look at the results reveals however that it is used in slightly different ways. In articles from the domain of hard news, it is often given as an absolute time rather than as a reference to a particular point in time, for instance when describing jail sentences. In articles from politics and economy on the other hand, units such as *financial year* or *fiscal year* are employed which are less relevant in the other domains. Apart from these specific usages, the collocations show that if *year* is used as a reference to a concrete point in time, it is most often to the past (*last year*, *previous year*, *past year, earlier this year*, *year ago*), while collocations like *next year* or *following year* are less frequent especially in lifestyle and hard news articles.

Looking at the remaining lexical items, the domains are clearly set apart. Every domain features words that reflect its specific thematic orientation, a distribution which can be found across the domains within the varieties as well (see appendix 2). The tendency of lifestyle and sports articles to feature more verbs among the most frequent lexical items than the other domains is particularly interesting; in the case of lifestyle articles, this can be seen as a reflection of the wide range of potential topics that are being treated here; there are no obvious nouns which define the subject as clearly as for instance *market* and *bank* for the domain of economy or *police* or *court* for hard news, which automatically renders the verbs more dominant in a frequency list. A look at the distribution in lifestyle articles within the varieties shows that this trend is general, and that the nouns which occur frequently are very broad terms such as *people*, *woman*, or *time*.

In the case of sports articles, the tendency can be explained with the more active nature of the subject of the domain. Especially the verbs *play* and *win* seem to reflect this. What is striking is that the verbs which are used are clearly oriented towards victory; *win* is among the most frequent items in sports news from all varieties, whereas its counterpart *lose* is never included even in the top 100. It produces just 51 hits in the Australian sports part (0.06% of the entire domain in this variety), 57 (0.06%) in the Hong Kong part, 63 (0.09%) in Kenya, 89 (0.08%) in the UK and 80 (0.07%) in US-American articles, and *defeat*, both as a verb and as a noun, is even rarer. Putting the focus on the winning side can thus be seen as a shared trend of sports articles across all five regions.

This observation is strengthened by an analysis not just of frequencies, but of keywords from the different domains and varieties.[22] A contrast of sports news from all regions with the articles from the other domains renders both *win* and *play* with high keyness values of 1,414 and 1,231 respectively. The strength of verbs in lifestyle articles on the other hand is relativised from this perspective, as the verbs frequently occur in the other domains as well. Table 5.2 summarises the top ten keywords for every domain, across all five varieties, in comparison with the other four domains.[23]

Table 5.2: Top ten keywords per domain

Economy							
No	Word	Freq.	Keyness	No	Word	Freq.	Keyness
1	market	1,294	2,718	6	its	1,866	1,565
2	per	1,471	2,058	7	shares	528	1,512
3	cent	1,291	2,026	8	investors	575	1,444
4	billion	1,000	1,998	9	stock	510	1,427
5	bank	833	1,732	10	growth	645	1,292
Hard News							
No	Word	Freq.	Keyness	No	Word	Freq.	Keyness
1	police	2,118	5,681	6	were	1,878	1,071
2	was	4,382	1,439	7	court	773	1,029
3	murder	422	1,249	8	said	4,040	972
4	crime	434	1,124	9	arrested	359	955
5	officers	405	1,094	10	killed	382	912
Lifestyle							
No	Word	Freq.	Keyness	No	Word	Freq.	Keyness
1	you	3,126	2,258	6	she	2,196	773
2	says	2,271	2,073	7	can	1,965	736
3	your	1,376	1,918	8	or	2,101	701
4	I	4,327	924	9	women	781	652
5	my	1,389	824	10	hair	306	587

[22] All keyword analyses for this investigation were conducted in AntConc (Anthony 2005) on the basis of word frequencies, using the log likelihood measure. For keyword analyses of the individual varieties, the data from the other varieties was used as a reference corpus, and the domains, too, were contrasted with each other. For variety- or domain-internal keyword analyses, data from this respective subset was used as reference, for example in a contrast of sports news from Australia versus the other domains from Australia (variety-internal) or political news from Hong Kong versus political news from the other regions (domain-internal). Of the results, only words with a keyness value of 15.13 ($p < 0.0001$) or more were considered key.
For the analyses, all words were treated independent of case and results were given out in lower case. In the following tables, proper names were changed to upper case again to increase readability.

[23] A more elaborate list with the top ten keywords for all domains within the individual varieties can be found in appendix 3. An earlier analysis of keywords in the sports part of the corpus can be found in Fest (forthcoming).

Politics							
No	Word	Freq.	Keyness	No	Word	Freq.	Keyness
1	election	1,061	2,871	6	president	887	1,311
2	minister	1,031	2,360	7	political	645	1,274
3	party	1,086	2,309	8	prime	574	1,270
4	Mr	2,343	1,993	9	coalition	363	950
5	government	1,480	1,520	10	labor	410	929
Sports							
No	Word	Freq.	Keyness	No	Word	Freq.	Keyness
1	team	1,339	2,749	6	coach	563	1,530
2	game	1,034	2,417	7	win	760	1,414
3	players	827	2,001	8	cup	514	1,356
4	season	843	1,910	9	play	681	1,231
5	league	665	1,604	10	match	497	1,105

As was the case with the most frequent lexical items, the keywords for the individual domains clearly indicate the dominant topics and in some cases also the dominant actors. As these keyword lists, in contrast to the list of most frequent items, were generated on the basis of words, not lemmata, the numbers represented in nouns and verbs elicit the subject matter further in this respect – in sports articles for instance *players* displays a higher keyness value than its singular form *player* (370 hits, keyness 898), and a similar trend is visible for *officers* in hard news and *investors* in economy. In political articles on the other hand *minister* and *president* are particularly key in their singular forms. Articles from the domain of lifestyle again stand out; of the ten most dominant keywords five are pronouns, leaving the potential actors undefined and more dependent on endophoric references. Another striking aspect in this context is the keyness of female actors represented by the words *she* and *women*. Within the domain of lifestyle, the two words do not hold particularly high relative frequencies (0.33% and 0.05% respectively), which demonstrates how subdued they are in the other domains to still reach such keyness values in the area of lifestyle.

Another notable difference between the domains is the degree of keyness that the keywords display. On average lifestyle has the lowest keyness values, which is certainly again due to its wide thematic scope and the resulting broad range of vocabulary. However, hard news also displays a very low keyness in general, with the huge exception of *police*, which has the highest keyness value across all lists and also clearly dominates this domain within the individual varieties (see appendix 3). The strength of this lexical item suggests that this might not be just a relevant topic, but that the involvement of police in a potential news story on the one hand raises its news worthiness in terms of impact and relevance (Bednarek and Caple 2014, 156) and thereby ensures its inclusion in a newspaper in the first place, and on the other hand is also a determinant for the story to be assigned to this domain. In this sense it would be an external criterion that is reflected one-to-one in the language.

When zooming into the domains and contrasting the parts from the different regions with each other (see appendix 3), the picture that emerges is first of all not very surprising. In all cases, place names display high keyness values, which is of course due to the fact that local news items have a higher relevance for the audience. In the case of national newspapers especially countries, counties, territories and major cities occur frequently and are therefore also keywords in their respective newspaper discourses. Least outstanding in this respect are the articles from the USA; here, US is a keyword in the domain economy only (in the form of U.S., which is why it results in two keywords *U* and *S*), in all other cases cities and, in the domain of sports, team names are more key. This can be traced back to the size of the country and the tendency to report on a more convenient scale of states or regions, but can also imply that the USA is simply mentioned more often in the other varieties as well since they functioned as the reference corpora for the keyword analysis.

Apart from place names, names of individual persons can be found especially in political articles, where presidents and other leaders are naturally relevant for the news. In articles from the UK the domain of sports news also reveals names of individuals, while in the other regions the sports itself is foregrounded. Here, differences can be made out with regard to which sports are popular in the respective regions – for Kenyan sports news, *marathon* occurs as a keyword, while news from Hong Kong features *sumo* and *soccer*. Sports articles from the USA do not present any particular sport among their top ten keywords, but instead work with team names such as *Knicks*, *Giants* and *Mets*, which every fan will recognise (Fest, forthcoming).

The contrast between the varieties and domains also shows some very fine-grained differences with regard to language use, such as the abbreviation *bn* for *billion* in the UK, *percent* instead of *per cent* or *%* in Hong Kong and the frequent inclusion of contractions with *'t* in the USA. Furthermore, the UK part displays an above average use of the first person singular pronoun *I*, especially in hard news, lifestyle and sports. Some more curious results can be found with regard to the address *Mr*; although it occurs more than once as a keyword for instance in Australia, it is the dominant one in the sports part from Hong Kong. All in all, it produces 303 hits in Hong Kong, 39 in the USA, 5 in Kenya and 4 each in Australia and the UK. The dominance in Hong Kong, however, can be traced back to one particular newspaper, the *Wall Street Journal Asia*, which is responsible for all hits in this variety. The American paper *Wall Street Journal* additionally contributes 37 of the 39 hits there, which makes the keyness of this lexical item very questionable (Fest, forthcoming). A summary of all variety-dependent keywords, always in comparison to the other four varieties, is given in Table 5.3:

Table 5.3: Top ten keywords per variety

Australia							
No	Word	Freq.	Keyness	No	Word	Freq.	Keyness
1	Australia	724	1,555	6	Abbott	255	687
2	Australian	633	1,482	7	WA	242	673
3	Queensland	234	806	8	Perth	197	641
4	Sydney	284	772	9	Brisbane	180	591
5	labor	359	704	10	Melbourne	246	589
Hong Kong							
No	Word	Freq.	Keyness	No	Word	Freq.	Keyness
1	China	1,660	3,605	6	Yuan	355	1,085
2	Hong	841	2,233	7	Beijing	407	1,025
3	Kong	840	2,202	8	Li	288	763
4	Chinese	791	1,675	9	mainland	264	740
5	HK	447	1,369	10	percent	427	642
Kenya							
No	Word	Freq.	Keyness	No	Word	Freq.	Keyness
1	Kenya	1,443	4,607	6	Kenyan	234	697
2	Sh	980	3,268	7	Mombasa	200	683
3	Nairobi	477	1,536	8	Kenyatta	235	657
4	county	590	1,284	9	Uhuru	193	638
5	Ruto	224	734	10	Kenyans	185	588
UK							
No	Word	Freq.	Keyness	No	Word	Freq.	Keyness
1	Britain	434	846	6	I	3,272	448
2	UK	402	734	7	bn	172	440
3	England	353	489	8	London	363	392
4	labour	261	457	9	Blair	148	385
5	British	364	449	10	Cameron	189	268
USA							
No	Word	Freq.	Keyness	No	Word	Freq.	Keyness
1	s	6,930	612	6	t	1,805	348
2	u	575	577	7	Berlusconi	148	270
3	York	365	502	8	Knicks	95	269
4	Boston	228	425	9	Weiner	79	239
5	Obama	308	367	10	center	218	230

Many aspects that occur in these keyword lists will be taken up again later in other operationalisations, such as pronominal use or the distribution of lexical items across and within word classes, which is why they will not all be discussed in detail here. One last aspect that should be noted however is the keyness of the words *mainland*, *China* and *Chinese* in all domains from Hong Kong. Although regional references are, as was said above, to be expected in newspapers, China is as dominant in the articles as Hong Kong itself, an observation which demonstrates what other studies have found already, namely the influence of self-censorship and adjustment to different political systems. Keywords and frequency lists alone cannot be a safe indicator here, yet in Hong Kong the tendency to cater for two regions appears much stronger than in the other four regions.

Tense

The second operationalisation to give an insight into the topic, or rather the temporal orientation of the topic, is that of the use of different tenses. Generally, news is about events that happened in the past (Harcup 2005, 107), yet a look at the distribution of tenses, depicted in Figure 5.1, shows that present tense is also used frequently.

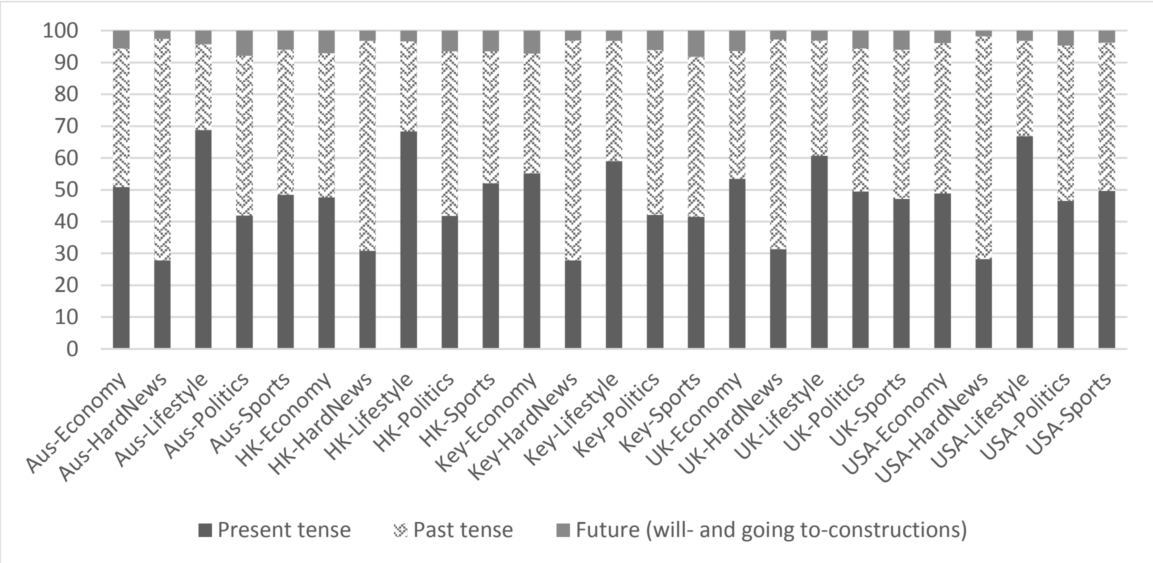


Figure 5.1: Relative distribution of tenses[24]

With regard to the distribution of present and past tense, huge differences can be found especially between the individual domains. The varieties all display a similar pattern: hard news contains the highest degree of past tense verbs in every variety, followed by politics (in Australia and Hong Kong), sports (in the UK and Kenya) or economy (in the USA), all three of which display similar values across all varieties as well and are clustered closely together. The domain of lifestyle stands out as contain-

[24] The calculation for Figure 5.1 is based on all finite verbs plus *will*- and *going to*-constructions. A more detailed discussion of tenses and expressions of time in the sports part of this corpus can be found in Fest (forthcoming).

ing by far the highest frequency of present tense verbs. The difference between lifestyle and the other domains is particularly expressed in Australia, where the difference to economy, which has the second highest present tense volume, is 17.92 percentage points, as well as in Hong Kong (16.22 percentage points) and the USA (17.19 percentage points), while in Kenya and the UK the divergence is smaller with 3.9 and 7.19 percentage points distance to the next domain respectively.

Across the five varieties, the domains therefore show parallel tendencies, but vary slightly in the degree to which these are expressed. Both in terms of these degrees and of ranking of the individual domains, the articles from Kenya show the strongest similarities to those from the UK. In both cases, differences between the domains are more moderate; between hard news with the highest frequency of past tense verbs and lifestyle with the lowest, the two varieties display just 31.35 (Kenya) and 29.73 (UK) percentage points difference. In contrast, in Australian articles the difference amounts to 42.64, in the USA to 40.1, and Hong Kong is the middle of the group with 37.77 percentage points.

Interestingly, these discrepancies in range between the two extremes are created almost entirely by the lifestyle domain. The past and present tense values of hard news as the one end of the scale are almost constant across the varieties, whereas lifestyle articles show a higher fluctuation. Comparing the domains taking into account each individual text, the median values confirm this observation (see Figure 5.2).

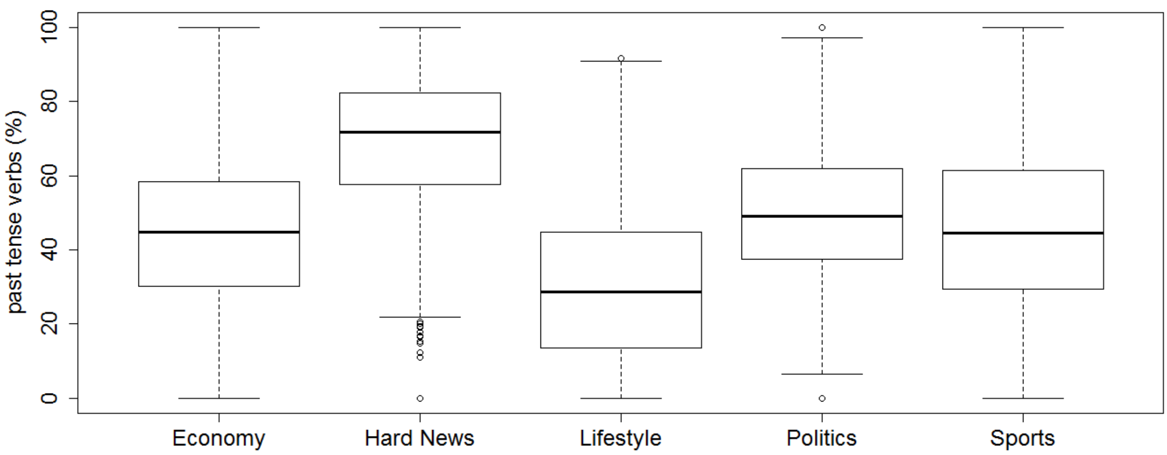


Figure 5.2: Distribution of past tense verbs within domains

In all domains except lifestyle, there are articles that contain only past tense verbs, even though for political news this value of 100% is considered an outlier. The medians reflect what has been discussed above, that the domains of economy, sports and politics display closely related patterns and distributions. Hard news shows the smallest range in terms of past tense usage across its articles – the median of the domain is by far the highest, and everything below 20% is displayed as outliers. Like the other domains hard news includes one article which contains no past tense verbs at all; this text is set apart clearly from the rest of the domain and a closer look at it reveals that it is a very short 71-word-long article from Australia which simply informs on a trial taking place on the day of the publication. It contains two verbs in present tense and two *will*-constructions. In general, the variation in terms of past and present tense usage between the varieties is highly significant at χ^2=11224.41 (df=4, p<0.000), to which the extreme tendencies in hard news and lifestyle articles contribute 6142.23 and 4767.34 respectively.

What can be concluded from these distributions are aspects of the temporal orientation of the individual domains. Hard news clearly focuses on the past more than articles from the other domains, while lifestyle constitutes the opposite. This underlines the assumption that lifestyle articles are not subject to the news value of recency as much as other news as the topics treated here are often not linked to a particular event in time but have a more general temporal scope. The dominance of personal pronouns as keywords in lifestyle news strengthens this further by indicating that the topics are more personal and geared more towards the readers themselves than towards passing on information about recent events, an aspect which will be analysed more closely in connection with the parameter of tenor.

The relative balance between the tenses that the domains of economy, politics and sports display does not allow any strong conclusions yet. Since the tagging of the corpus does not include the option to filter out quotations, it has to be taken into account that direct quotes are likely to include present tense verbs which do not necessarily have an impact on the general orientation of the text (see Fest, forthcoming). To some degree, this holds true also for headlines; as was described in section 4.2.2, present tense is a characteristic of this part of newspaper articles, and indeed 3.38% of all present tense verbs occur in first or second headlines. These do not amount to many instances per text however, as the amount of headlines is naturally limited, and a control query showed that the overall distribution remains the same. Furthermore, there are also 815 hits of past tense verbs in the headlines, which means that although the present tense is stronger in this part (the difference to the cumulative main bodies of the texts is significant at χ^2=154.64, df=1, p<0.000), it is not used exclusively. In contrast to direct quotes headlines thus appear to have the potential to contribute to the temporal orientation of the text, which is why verbs used in headlines were included in the calculations above.

One aspect that was so far left aside is that of future references. The first very obvious conclusion to be drawn on the basis of the distributions shown in Figure 5.1 is that future references in the form of *will-* or *going to*-constructions are very rare. Of these two, *will*-constructions are still by far the more frequent ones; a calculation on the basis of the individual texts renders *going to*-constructions as almost irrelevant, resulting in a median value of 0 for every category. This does not necessarily mean that the articles do not include any references to the future, but implies that potential references are realised not via tense, but by employing for instance adverbs of time. A closer analysis of such lexical items is therefore necessary to include the future aspect properly and not draw false conclusions. Furthermore, even the more concrete use of past and present tense determines the overall orientation, but not the scope of the time that is referred to. The next operationalisation, temporal expressions, should therefore be considered in close connection to the distribution of tenses.

Expressions of time

Expressions of time are of course not a closed semantic class. Temporal relations can be expressed in manifold ways and are often dependent on the context in which they appear. For the purpose of a quantitative analysis, a concrete list of items is therefore difficult to define, yet a number of suggestions can be found in the literature.

First of all, a lot of temporal references are constituted by adjectives and adverbs, and Biber et al. (1999, 552-53) distinguish between four ways in which an adverb can express meanings of time, namely in terms of position, frequency, duration, and relationship. For this investigation, those adverbs that focus on a position in time are most relevant, as they most clearly reflect the concrete times to which an article refers and thereby determine the scope of the news story. Of the adverbs listed by Biber et al. (ibid., 561) as the most common ones, we will therefore include *now*, *today*, *ago*, and *yesterday*. Furthermore, *tomorrow* will be added as the counterpart of *yesterday* and, in addition to *ago*, which only receives its full meaning by its collocates, *last*, *previous*, and *next* will also be included.

For these last four, collocation lists were generated beforehand to distinguish the relevant temporal uses from other constructions. Based on these, the most frequent combinations of the respective keyword and lemmata with a temporal meaning will be included here:

last + [*year/week/month/night/season*]
next + [*year/month/week/season*]
previous +[*year/month/week*]
[*year/month/week/decade/day*] + ***ago***

Of course, all of these constructions can be used in different ways; they can rely on the time of the article as a reference point, which is what we are interested in here, but also draw on some other point in time previously defined in the article. *Previous year*, for instance, can refer to the year previous to now, but also previous to some other reference point that can be further in the past, or also in the future, depending on how the journalist structures the article in question. Since an analysis of these differences would require a manual examination of every instance of these terms, for the purpose of this investigation they will be generalised as references to the past, present or future respectively.

First of all, the lexical items under examination reflect slightly different scopes. *Yesterday*, *today* and *tomorrow* refer to the immediate time frame, while the most frequent collocates listed above, *year*, *month* and *week*, imply larger temporal distances. Within the narrower scope, the past and present are generally the dominant reference points, as Figure 5.3 shows.

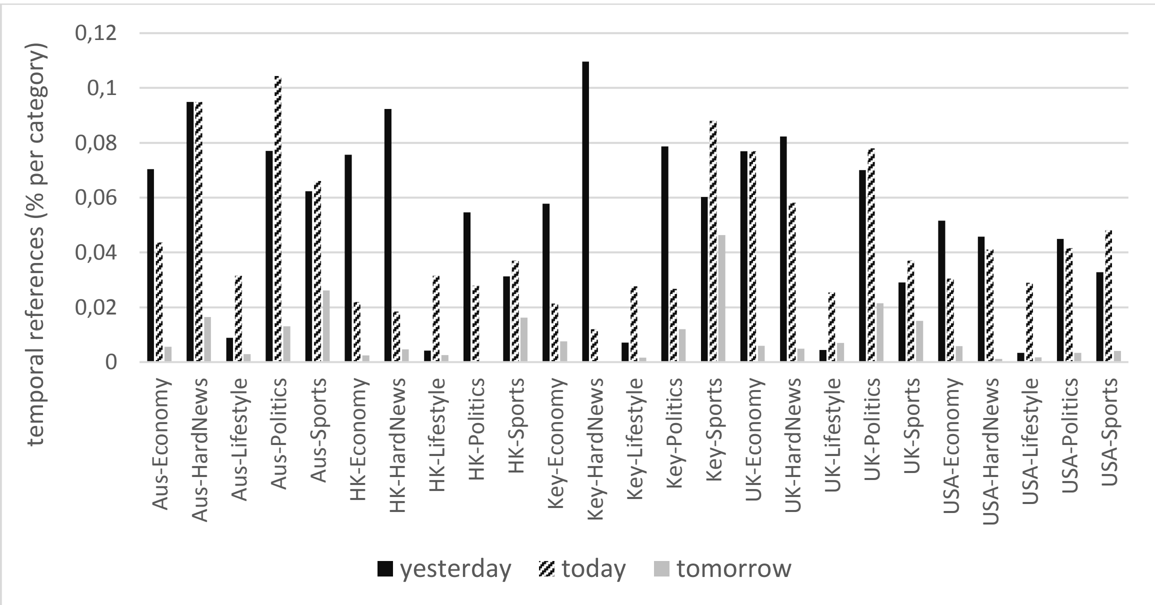


Figure 5.3: Relative frequencies of temporal references with a narrow scope

Despite this general trend, some clear differences can be observed across the domains as well as the varieties.

Regarding the domains, lifestyle again displays a particularity as it contains the fewest temporal references in the form of *yesterday*, *today* or *tomorrow* in every variety. Furthermore, of the references included in this domain by far the largest number refers to the present, which confirms the observations that were made regarding the distribution of tenses, that present tense was the dominant form in lifestyle articles. Sports news, too, displays a general trend featuring *today* as the dominant reference in all varieties. Additionally, it always contains the highest or second highest amount of *tomorrow* (cf. Fest, forthcoming).

The other domains are less congruent across the varieties; hard news tends to prefer references to the past and contain hardly any to the future, which is expressed particularly clearly in the hard news from Kenya and Hong Kong. In Australia on the other hand, present references equal past references in terms of frequency. This Australian particularity is mirrored also in political news, where *today* occurs more frequently than *yesterday*. This is true also for political news from the UK, while in the other varieties they are oriented more towards the previous day.

These distributions across the temporal scale immediately surrounding the article, from the day before to the day after, allow some preliminary inferences regarding the topics of the domains. In sports news, where references to the present are dominant and outlooks to *tomorrow* are more frequent than in the other domains, we can assume that for instance articles containing match reports are written on the same day as the match, making *today* a reference to the very close past (e.g. *earlier today*) and that the news includes announcements of issues like player transfers or match previews which are to happen in the course of the day. Once again, direct quotations have to be taken into account as a source for *today*, as they can include this reference without changing the scope of the article itself (cf. ibid.).

For lifestyle articles, the generally low frequency of immediate temporal references again indicates that recency is not the major news value constructed in these articles. The relative dominance of *today* is striking within the domain, yet the overall amount is so low that a well-founded interpretation is hardly possible. For hard news on the other hand, the clear dominance of the reference *yesterday* shows that the temporal dimension is of high relevance here and that stories included in this domain are not just reported right on the next day, but that the assignment of the time of the event seems to be important as well. This is also reflected in the total numbers of instances of *yesterday*, *today* and *tomorrow*, which are highest in the hard news domain.

In addition to these three temporally narrow items, the expressions *last*, *previous*, *next* and *ago* are more flexible in their range as they depend on their collocates. Nevertheless, they can be assigned to a direction they refer to. Figure 5.4 shows the distribution of references to the past, present and future across all categories.

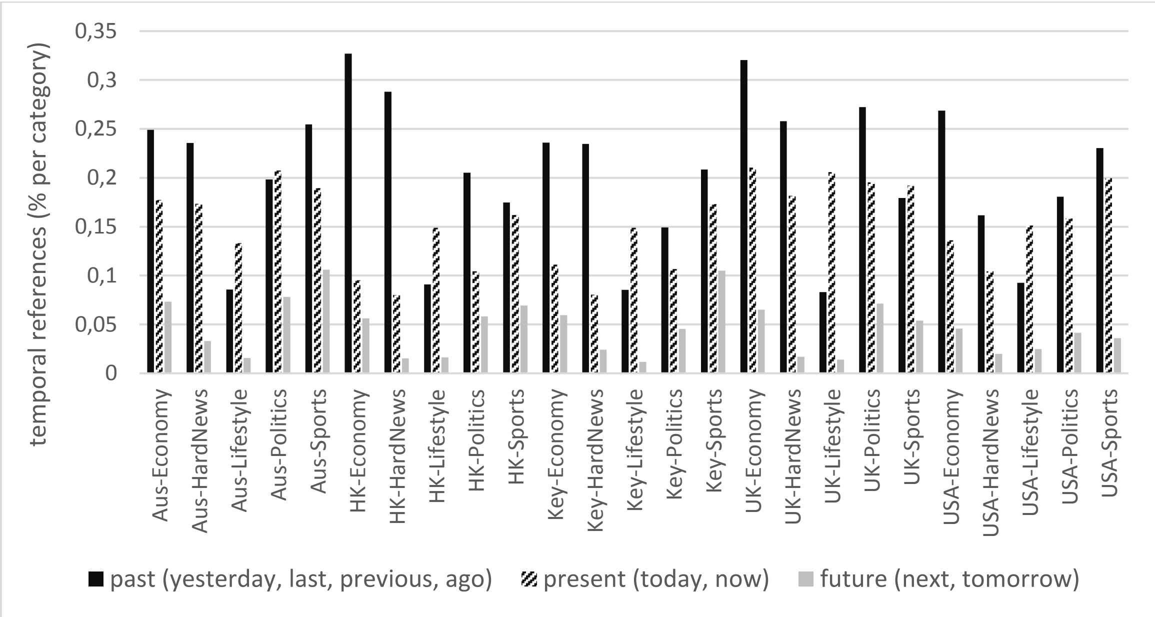


Figure 5.4: Relative frequencies of temporal references

When including the more flexible items of reference, the picture changes slightly in some aspects. For lifestyle articles, the trend that present references are dominant remains the same, and for hard news, too, the tendency towards past references is confirmed. However, hard news no longer contains the highest overall amount of temporal references; they only make up 1.94% of all word tokens, while for news from the domain of economy, they amount to 2.1%. Economic news also displays a clear orientation towards the past now, and political and sports news follow this trend, although not as strictly. Especially for sports news, the references to the future also increase when including *next*, which can be traced back to the collocation with *season*. While the other domains mainly use *year*, *month* or *week*, as the list of collocates shows, *season* has a very specific meaning in many areas of sports and defines as concrete a time span as *year*. In the other domains, such expressions occur (e.g. *fiscal year* or *financial year*), but they are much rarer.

Since *previous, last, next* and *ago* are in themselves independent of an exact time span or position, their use alone only shows a direction but gives no indication about the scope implied. The major collocates of all four are the lemmata *year*, *month* and *week*, the frequencies of which, as collocates of one of the four, are shown in Figure 5.5.

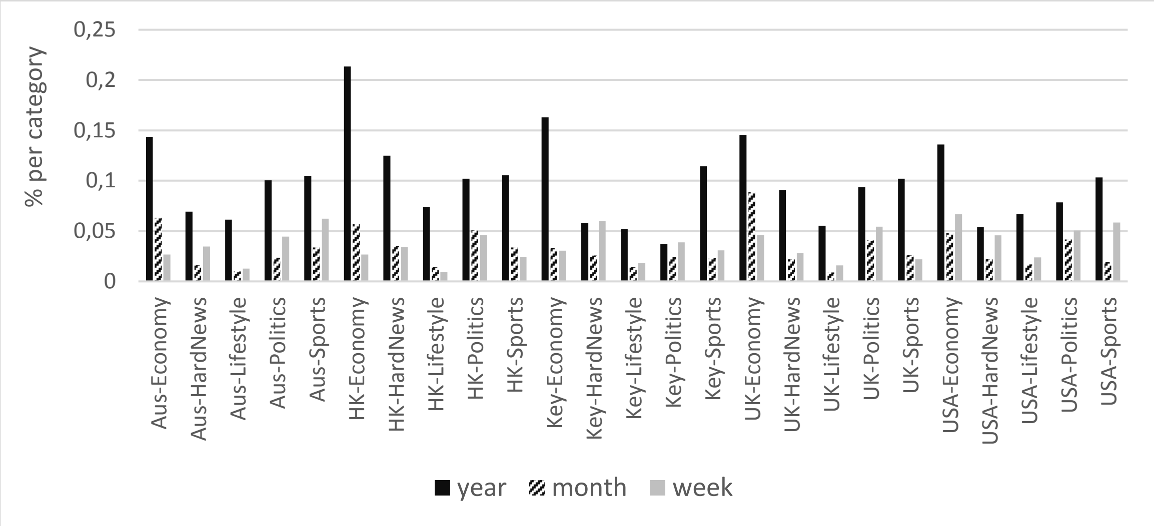


Figure 5.5: Relative frequencies of [last|previous|next|ago] + year, month, week

With the exceptions of hard news and politics in Kenya, the lemma *year* is the most frequent item of the three. This is not surprising as it already occurred in the keyword lists discussed above. In most cases, *week* slightly outweighs *month* as a reference scope.

When looking at the concrete time spans that are connected to these lemmata, it becomes clear that in the vast majority of cases, they refer to one month, year or week. In the other cases, *year* as a reference is used either in the frame of two or three years, or in combination with round numbers like 10, 20, or 100. *Week* often co-occurs with two, thus referring to a fortnight, and *month* often defines a scope within 6 months, although *12 months* as a lexical alternative for *year* also occurs. Another lexical combination which has been left out so far is *this + year|month|week*, as it is very difficult to define the temporal direction of this construction. What can be said for certain is that it puts the emphasis on the relation to the present time instead of on how far away something is; *this week* for instance constructs the news value of recency much more effectively than *six days ago* does. It is therefore not surprising that *this* occurs most frequently with *year*, as a story that happened several months ago sounds old and decreases in relevance, whereas the inclusion of *this* puts it into perspective to the current time.

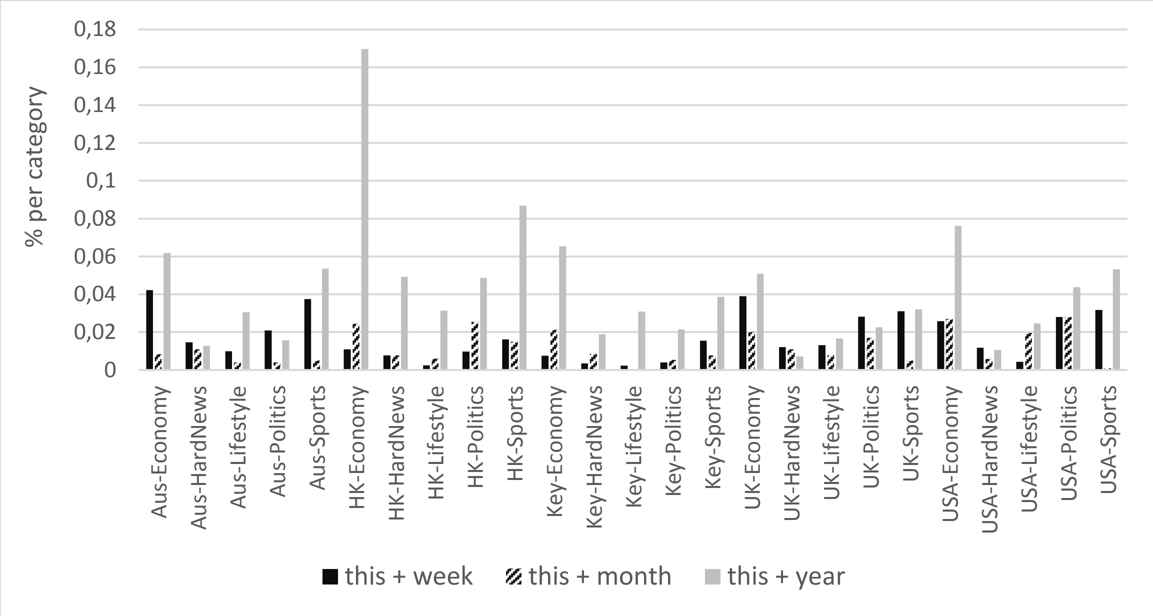


Figure 5.6: Relative frequencies of this + [year|month|week]

In contrast to the distribution of *year*, *month* and *week* in general as shown in Figure 5.5, these more concrete constructions display an almost even relevance of *month* and *week*. A characteristic can be found for sports news, for which *week* is clearly surpassing *month* in all varieties, which can be attributed to the structure of many league games to have weekly match days.

To briefly sum up the most important aspects of temporal orientation and scope in the articles in this corpus, it can be said that in general, references to the past are most dominant, whereas outlooks to the future are least frequent. The scope is most commonly defined in a unit of one (year, month or week), which can be attributed to the same desired effect as the frequent use of *this*, namely the increase of recency and thereby relevance.

In terms of variation, the varieties show less discrepancies than the domains. Especially lifestyle articles stand out as containing the fewest references; the items analysed here amount to only 1.31% of the entire lifestyle-part, while economy (2.31%), hard news (2.07%), politics (1.82%) and sports news (2.06%) produce much higher values. In terms of internal distributions, the domains show similar patterns in Kenya and Hong Kong and are closest to that observable for the USA. These three varieties also display the lowest total values of references with 1.61%, 1.8% and 1.87% respectively.

Lexical density

The last operationalisation of the experiential domain is more general in its results and does not reflect the concrete content as much as the features discussed above. Lexical density is a measurement of the relation of content and function words and therefore serves as an indicator of the amount of information that is contained in a text. It displays a text's percentage of content words, which have been defined here as nouns, verbs, adjectives, proper nouns and content adverbs (tagged as RB, RBR and RBS, see section 4.2.2).

As the corpus was normalised on the basis of articles, not words, a first look should be taken at the variation in terms of article length. This differs greatly especially between the domains, but also between the varieties, and can therefore be regarded as a distinctive, although not purely linguistic, feature between the categories (see Figure 5.7)

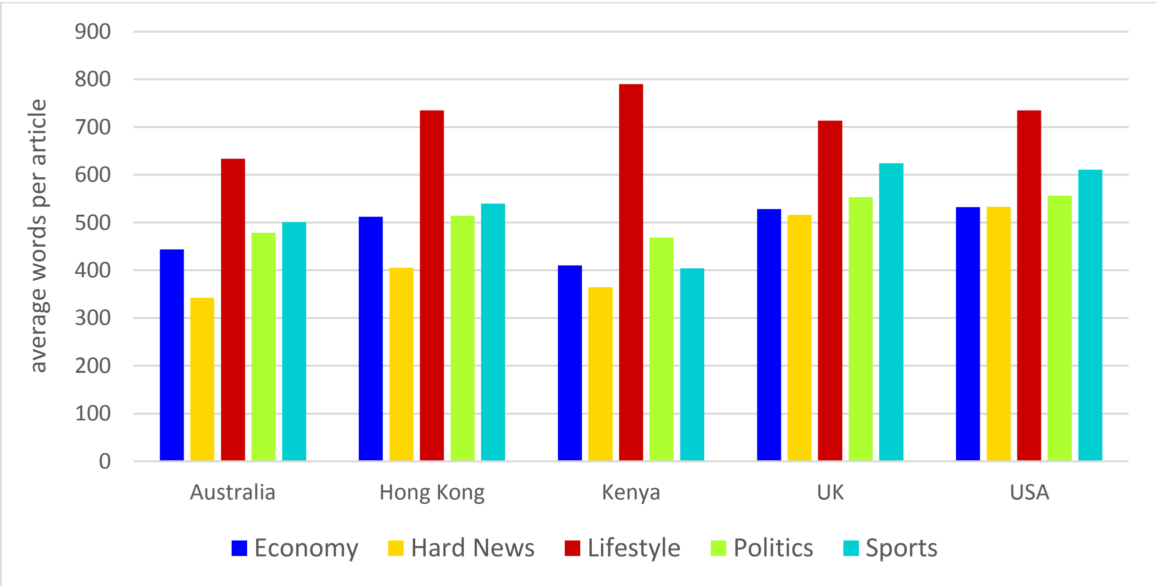


Figure 5.7: Length of articles in number of word tokens

The bars in this chart represent the average article length per category. As can be seen, the domain of lifestyle contains the longest articles, while hard news items are the shortest in all varieties except the USA, where economy has a slightly lower average (and indeed contains 59 words less than the hard news sample).

In terms of lexical density, hard news, economy and politics display the most constant values across the varieties. In all three cases, the news from the UK produces the lowest density values while Hong Kong and the USA show the highest. For the domains of lifestyle and sports news, a lot more variation is observable, as Figure 5.8 shows.

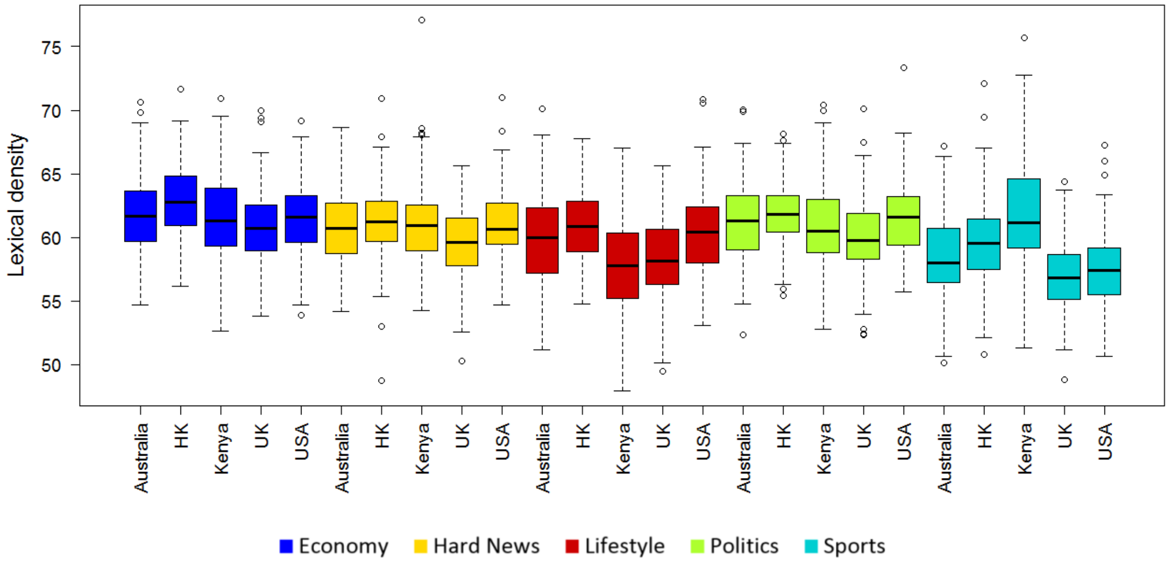


Figure 5.8: Variation in lexical density

With regard to lifestyle articles, news items from Kenya show the lowest median for the density value and with this also display the second lowest value of all categories, after sports news from the UK. The Kenyan part is also responsible for the biggest range between highest and lowest density among lifestyle articles, but the distribution is even as it produces no outliers. This appears characteristic of lifestyle articles in general; in contrast to all other domains the categories here show hardly any outliers and also feature relatively centrally positioned medians, which represents a very even, though widespread, distribution of density values.

The most outstanding domain among the five however is sports. While density values here appear quite compact and similar in the USA and the UK, the Hong Kong part also shows a rather small range of values, but a higher median than the two. Kenya once again displays the highest range, and at the same time produces only one outlier, which is less than any other variety. This pattern of an even distribution is constant for Kenyan news, with the exception of hard news, where it produces more outliers than the other varieties. This however is due to one extreme value which influences the calculation, and an examination of this article reveals that it is a very short item of just 48 word tokens. This of course means that any potential pronominal references are superfluous and conjunctions are hardly necessary, as the article consists of just three sentences and the headline, and thus automatically reaches a higher lexical density.

The measurement of lexical density is generally slightly dependent on the length of the text, which is problematic in newspaper discourse which, in this collection, contains an average of 538 word tokens per text. For the current dataset, a Pearson's correlation test reveals a correlation of r=-0.28, which indicates a weak negative correlation of the two values. Although the dependency is thus not overly strong, it explains the high density values of outliers like the Kenyan hard news article and should be kept in mind with regard to extreme values in either variable.

In general, the conclusions to be drawn from the distributions of density values shown above are rather broad. This is sufficient for an analysis of the experiential domain in combination with the other operationalisations, yet for later purposes, the distribution of individual word classes will be more relevant. For the time being, a number of results can be summarised; on the one hand, Kenyan news displays the widest, and yet most even range of density values. News from the UK has the lowest average density value and also the lowest averages in every domain except lifestyle. Among the domains, sports articles stand out as displaying the lowest lexical density on average, which is not quite in line with the results from the keyword analyses in which the lifestyle domain rendered many pronouns. The difference between these two domains in terms of average density is small however (0.53 percentage points; 59.44% for lifestyle, 58.91% for sports). Economy, hard news and politics also differ only slightly amongst each other, displaying averages of 61.7%, 60.69% and 61.17% respectively.

With regard to the content, this indicates not only that lifestyle and sports articles have lower information values, but also that these articles are more focused on one topic or person. If only few actors appear in a story, personal references can be included more frequently as the risk of ambiguity is lower and fewer people need to be introduced. The same holds true for place references or team names. Furthermore, these two domains might make less use of the strategy to replace personal pronouns by pieces of information as shown in example [1] (see section 3.2.1.2):

[1] **Heather Watson** is hoping glandular fever will not rule her out of next month's French Open. [...] **The 20-year-old** has not picked up a racket since her diagnosis three weeks ago and says she has no idea when she will be well enough to return to the practice court. (Hart 2013, *The Daily Telegraph*; emphasis added)

This in turn is all the more likely if 1st or 2nd person pronouns are frequent; if the reader is addressed directly or the author speaks about him- or herself, these substitutions rather appear in addition to the pronoun, but not instead of it, as the following excerpt exemplifies:

> [2] Why is there sugar in all these things? Partly, it's because **we**, **the consumers**, want it. We want everything to be sweeter these days; we wouldn't like to go back to the good old days, when smoked salmon tasted fishy and salty and dried cranberries tasted tart. (Leith 2012, *The Daily Telegraph*; emphasis added)

Leaving out the pronoun *we* in this context would turn this into a 3rd person reference to consumers in general, while the inclusion of *we* ensures that the readers are addressed and the author is involved as well. For such purposes, the pronoun is therefore necessary, and neither pieces of information nor proper names can substitute it without losing the meaning, they can only occur additionally. As pronouns are counted as function words, they lower the lexical density value. Texts that aim at more personal dimensions are therefore likely to tend towards a lower lexical density, which is confirmed in the case of lifestyle by the pronominal keywords. For sports news, pronouns are among the top ten keywords in Australia, the UK and the USA (see appendix 3), although they are by far not as dominant as in lifestyle articles. Since this is already within the range of social relationships between the discourse participants, a closer look at different pronouns will be discussed under the parameter of tenor in section 5.2.

5.1.2 Goal Orientation

To complete the subject matter of the discourse, goal orientation will be analysed as the second subdimension of field. As was discussed in section 3.2.1.2, this is problematic on text level and will therefore be considered here only in terms of informativeness. For this purpose, the three variables of type-token-ratio, mood and pronominal use were chosen as operationalisations. Especially for the last two, in-depth analyses are parts of other register parameters, which is why goal orientation will be concerned mainly with quantitative distributions.

Type-Token-Ratio

The type-token-ratio is a measurement of the variation within the vocabulary of a text. It represents all tokens in comparison to the different lemmata that a text includes. In general, there is a high correlation between the type-token-ratio and the article length (r=-0.76 for the corpus at hand). For the type-token-ratios for individual word classes, the dependency on the overall length is a little lower (r=-0.56 for verbs, r=-0.37 for nouns, r=-0.29 for proper nouns, r=-0.39 for adjectives, r=-0.43 for adverbs). Instead, they display high correlations with the total number of tokens within the respective word classes. If a text for instance includes only four adjectives, two of which are the same, it has three different types and the type-token-ratio will be 3/4, or 75%. This correlation cannot be avoided, which is why an analysis of the range of this value across all texts, as represented in a boxplot, is rather misleading than helpful, and word class-internal measurements were found to essentially represent the distribution of absolute values. As Figure 5.7 showed, the absolute article length varies greatly across the corpus; within the individual categories, the range is smaller, which is why an average on this level appears most accurate in terms of type-token-measurement. Table 5.4 presents these values taking into account all parts of speech.

Table 5.4: Average type-token-ratio across domains and varieties

	AUS	**HK**	**KEY**	**UK**	**USA**	*whole domain*	*range*
Economy	44.40	42.30	43.13	43.92	42.55	3.19	2.1
Hard News	46.49	45.77	45.39	42.47	43.06	3.67	4.02
Lifestyle	40.98	40.26	37.69	40.19	39.46	3.33	3.29
Politics	42.88	43.19	41.87	43.98	45.16	3.22	3.29
Sports	42.61	42.79	43.39	41.65	39.08	3.15	4.31
whole variety	3.66	3.56	3.18	3.42	3.56	1.47	
range	5.51	5.51	7.7	3.79	6.08		8.8

With regard to the domains, the first observation to be made is that lifestyle again stands out. It has the lowest average type-token-ratio in every variety except the USA, where sports news is slightly lower. Especially in Kenya, the low type-token-ratio of the lifestyle domain is responsible for the comparatively wide range in the variety, as its highest value, displayed for hard news, is not above the highest values of other regions.

Hard news, too, are once again set apart from the other domains by holding the highest value on a total average, although in the USA and more so in the UK it does not hold this position. Their generally high value can be explained by the fact that the articles in this domain are also on average the shortest, but also by the relevance of places and people in these stories. The keyword analysis showed that the strikingly dominant actor in hard news stories is the police, which indicates that topics to a large degree consist of accidents and crimes. In these contexts, the six *W*-questions that guide a reporter – *Why? What? When? Where? Who? How?* – are of particular interest and constantly require new types in the vocabulary. The replacement of pronouns by bits of information offers itself here, which additionally increases the type-token-ratio as pronouns per se have less potential for variation.

As the example of hard news shows however, type-token-ratio cannot necessarily be equated with an abundance of topics or even a wider focus within one topic. Even a story that includes very few actors, like a car accident, can increase its type-token-ratio simply by answering the six *W*-questions without shifting the focus from the accident or the people involved, or without intending to comment on any of the facts further. For the purpose type-token-ratio is supposed to fulfil here, namely a measurement of informativeness, these aspects are not relevant, yet it is important to note that interpretations of this measurement in these directions can only be tentative.

With regard to the varieties, it can be seen that the UK displays the most compact picture. With a range of just 3.79% the type-token-ratio is relatively constant across the domains, although within hard news, the UK has the lowest value. Kenya and the USA are to be found on the other end of this scale, with comparatively large internal ranges. For the informativeness the values point towards more diversity amongst the domains than amongst the varieties, which is in line with the results of the previous linguistic features.

Mood

In contrast to the type-token-ratio, which is a rather crude measurement due its dependency on the absolute length of the text, which mood the author chooses to use is independent of the amount of tokens or words. Mood is included in this study twice, as an operationalisation for both goal orientation and social role relationship, but will be analysed with different foci. In the context of goal orientation, the quantitative distribution is of major interest, as the degree of informativeness is taken to be represented primarily by declaratives. The analysis at this point will therefore be kept rather brief and will not include any in-depth investigations of interrogatives or imperatives. In the framework of social hierarchy, this in turn will be the focus, as the implications of these two moods are highly relevant there.

For the quantification of mood, we distinguish between interrogatives, imperatives and declaratives. Since the corpus does not contain parsing or other annotation for phrase analysis, the counts rely on the possibilities offered by the part of speech-tags. For the queries, the punctuation marks *. ! ; : ?* were taken into account. Commas were left out, as large portions of the hits the respective queries produced were irrelevant. Imperatives were defined as base verb forms in sentence-initial position and were queried with the following command:

> [word="\.|!|;|:|\?"][pos="''|``"]?[(pos="VB")&(lemma!="have|do")];

The first element defines the string to start with the end of a sentence, the second, optional element allows for the possibility of a quotation. The third element determines the part of speech-tag to be VB for a verb in its base form and at the same time excludes the lemmata *have* and *do* at this point. This was included as these two verbs in sentence-initial position are characteristic of questions. The instances were queried separately and resulted in 38 hits, of which the ones functioning as imperatives were filtered manually. Interrogatives on the other hand could be queried more simply using

> [word="\?"];

Declaratives, then, were all other sentences. As was to be expected, this last category of mood is clearly dominant across all varieties and domains, yet some variation can be observed (see Figure 5.9).

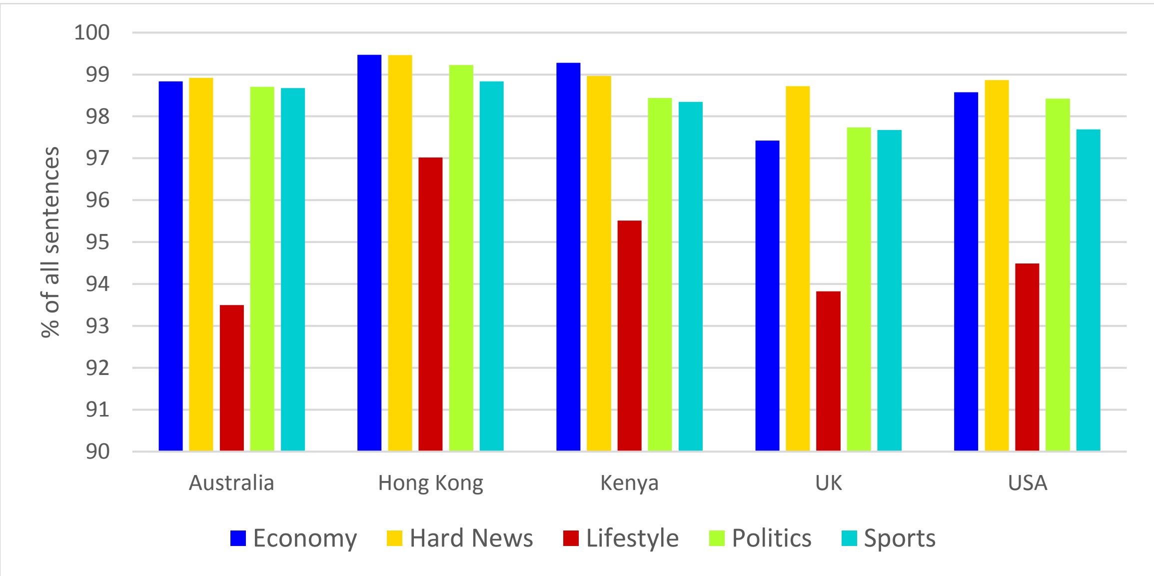


Figure 5.9: Relative frequencies of declarative mood

Regarding the varieties, the variation between the five regions is significant at χ^2=141.63 (df=4, p<0.000), which is mainly due to the dominance of declaratives in

Hong Kong and the rather low values in the UK. The domains display an even stronger level of significance of their differences, at χ^2=1176.4 (df=4, p<0.000). In all varieties, the domains of economy, hard news, politics and sports are close to each other though, while lifestyle has much lower values. This becomes even more evident when looking at the distribution across the individual texts, as depicted in Figure 5.10.

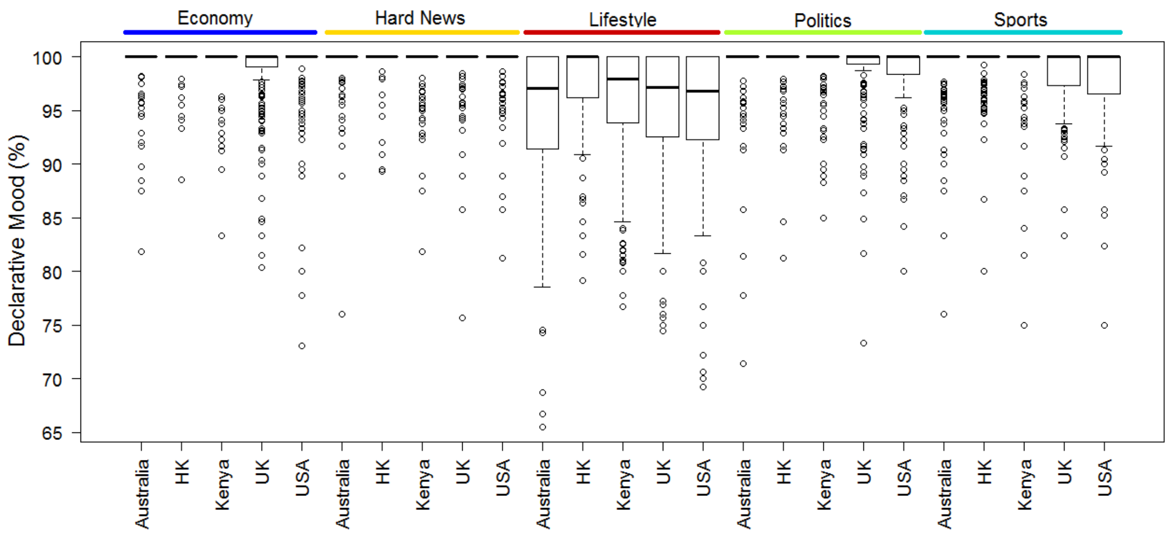


Figure 5.10: Distribution of declarative mood

As many texts do not include any imperatives or interrogatives, the median values for declarative mood in most categories are naturally high. In hard news, there is no exception from the median equalling 100%, although a number of outliers represent individual irregularities in this domain. In economy, the comparatively low value from the UK part in terms of averages is shown to be not just the product of one or two extreme values, but depicts a slightly broader variation despite numerous outliers. For politics and sports, news items from the USA share this trend and especially regarding sports produce only few outliers, indicating a more even distribution of the values. The wide range of lifestyle of course means that here, imperatives and interrogatives are more common, which will be analysed in more detail for the subdimension of social role relationship.

In terms of the degree of informativeness, we can therefore assume that there is little variation between the domains of hard news, economy and politics, and that even sports, which stood out slightly in the other analyses, is closer to these three domains than to news from lifestyle. The values for lifestyle news on the other hand further confirm the assumption that the topics here have an interpersonal focus which is expressed, among other things, in the usage of imperatives and interrogatives.

Pronominal use

As was described in section 3.2.1.2, informativeness in news language in terms of pronominal use cannot be restricted to the 3rd person singular *it*, but has to take into account references to persons as well since their involvement alone can be the dominant factor of news worthiness. In this sense, the other 3rd person pronouns are of particular interest for the subdimension of goal orientation as well. All pronouns were queried including their possessive forms, which led to the final list consisting of *he, his, him, himself, she, her, herself, it, its, itself, they, their, them* and *themselves*.

Across the whole corpus, the trend that differences are larger between the domains than between the varieties is observable also for this linguistic feature, although it is not expressed as strongly. Table 5.5 shows the relative frequencies of 3rd person pronouns, of all pronouns, and the percentages of 3rd person with relation to all pronouns.

Table 5.5: Relative frequencies of 3rd person and all pronouns

		AUS	HK	KEY	UK	USA	*domain*	*range*
ECO	3rd p. pron.	2.57	2.03	2.04	2.81	2.49	2.4	0.78
	All pron.	3.43	2.37	2.76	3.61	3.13	3.07	1.24
	3rd p./all	74.93	85.65	73.91	77.84	79.55	78.18	11.74
HN	3rd p. pron.	4.18	2.78	3.92	5.2	3.98	4.07	2.42
	All pron.	5.17	3.17	4.71	6.53	4.95	4.99	3.36
	3rd p./all	80.85	87.70	83.23	79.63	80.40	81.56	8.07
LIFE	3rd p. pron.	3.62	3.27	4.4	4.22	3.44	3.8	1.13
	All pron.	6.74	4.92	7.06	7.42	5.72	6.37	2.5
	3rd p./all	53.71	66.46	62.32	56.87	60.14	59.65	12.75
POL	3rd p. pron.	3.2	2.81	3.29	3.78	3.39	3.3	0.97
	All pron.	4.41	3.31	4.17	5.15	4.35	4.29	1.84
	3rd p./all	72.56	84.89	78.90	73.40	77.93	76.92	12.33
SPO	3rd p. pron.	4.19	3.49	3.17	4.67	4.36	4.04	1.5
	All pron.	6.61	4.94	4.77	7.25	6.49	6.12	2.48
	3rd p./all	63.39	70.65	66.46	64.41	67.18	66.01	7.26
variety	3rd p. pron.	3.54	2.92	3.51	4.15	3.55	3.54	
	All pron.	5.41	3.87	5.05	6.11	5.02	5.11	
	3rd p./all	65.43	75.45	69.50	67.92	70.72	69.44	
range	3rd p. pron.	1.62	1.46	2.36	2.39	1.87		3.17
	All pron.	3.31	2.57	4.3	3.81	3.36		5.05
	3rd p./all	27.14	21.24	20.91	22.76	20.26		33.99

Looking only at 3rd person pronouns, among the varieties, the lowest average is displayed by news from Hong Kong, while the UK part contains the most. In Hong Kong news especially the domain of economy stands out as containing the fewest pronouns across the whole dataset; the trend for this domain to be at the lower end is constant across all varieties however, and it shows little internal variation. The same holds true for politics, which takes different places within the individual varieties, but contains a relatively small range of percentages in itself.

The other three domains are more subject to variation. Hard news contributes the highest values in the UK and the USA, while they are less extreme in the other varieties. The most dominant pronoun in this domain are all forms of the singular *he* (6,471 occurrences, 1.87%), followed by *they* (2,918 occurrences, 0.84%). With this, it is similar to sports news, where the dominance of *he* is even slightly stronger with 8,463 occurrences (1.97% of sports news in total, 48.76% of 3rd person pronouns in sports news). This aspect sets the two domains apart from the area of lifestyle, where *it* is most dominant (6,546; 1.13%) before *they* (6,052; 1.05%) and *he* (5,321; 0.92%).

At a first glance, these results seem to contradict the assumptions made above that lifestyle articles focus on people and human interest aspects more than hard news does; however, a look at all pronouns puts the 3rd person instances into perspective. Hard news only contains 4.99% pronouns in general, the 3rd person pronouns with 4.07% are therefore the major group. Lifestyle articles on the other hand contain 6.37% in total, of which only 3.8% are contributed by 3rd person ones. This observation on the level of domains across varieties is also visible within the regions; the smallest discrepancies between all pronouns and 3rd person forms are to be found in the domains of economy, politics and hard news, while lifestyle and sports articles display the largest differences.

In this context, the results rather confirm what was suggested by the indicator of mood, that interpersonal dimensions are particularly relevant in the two domains. Looking at the distribution of the 3rd person pronouns and its meaning for the degree of informativeness, the results point towards a highly informative structure in economic, political and hard news, and indicate that among the varieties, especially Hong Kong puts the focus on information, as the variety shows the clearest dominance of 3rd person pronouns in all domains.

5.1.3 Summary

Summing up the results for the parameter of field, it can be said that in general, there are more differences with regard to functional than to regional variation. From among the domains, economy and politics often display similar values, while the other three show particularities in at least one linguistic feature. The keywords clearly represent the topics that are dominant in the respective fields, and also reflect the regions by including references to places and important actors. Lifestyle and sports articles present strong keyness values for verbs and pronouns, while the only variety set apart is the USA with fewer key place references.

With regard to temporal orientation and scope, a general tendency can be observed to refer to the past, but stress the recency of the news story. This is particularly true for constructions with the lemma *year*, which is most often combined with *this*. Lifestyle articles display the fewest temporal references of all domains and thereby put the weakest emphasis on constructing recency. If at all, articles from this domain tend to concentrate on the present. This is true also for sports news, which contain comparatively frequent future references as well. The opposite is reflected in hard news, where the dominant reference point is the past with a strong focus on the immediately previous day. The consistency of economy, hard news and politics in contrast to the more diverse pictures of sports and lifestyle is reflected also in the lexical density. While the first three domains are again clustered fairly closely together, both sports and lifestyle news display broader internal ranges as well as few outliers, indicating an even distribution across the individual texts. Variation in this aspect therefore appears to be more normal, whereas in economy, politics and hard news, there is a narrow standard that is breached only by few texts with extreme values.

As the lexical analyses of frequent items and keywords were qualitative, a quantitative representation of this linguistic feature is not possible at this point. Since the keywords were well distinguishable however, this is not necessary. The other two aspects, temporal orientation and lexical density, can be put into relation quantitatively, but produce only vague patterns (see Figure 5.11).

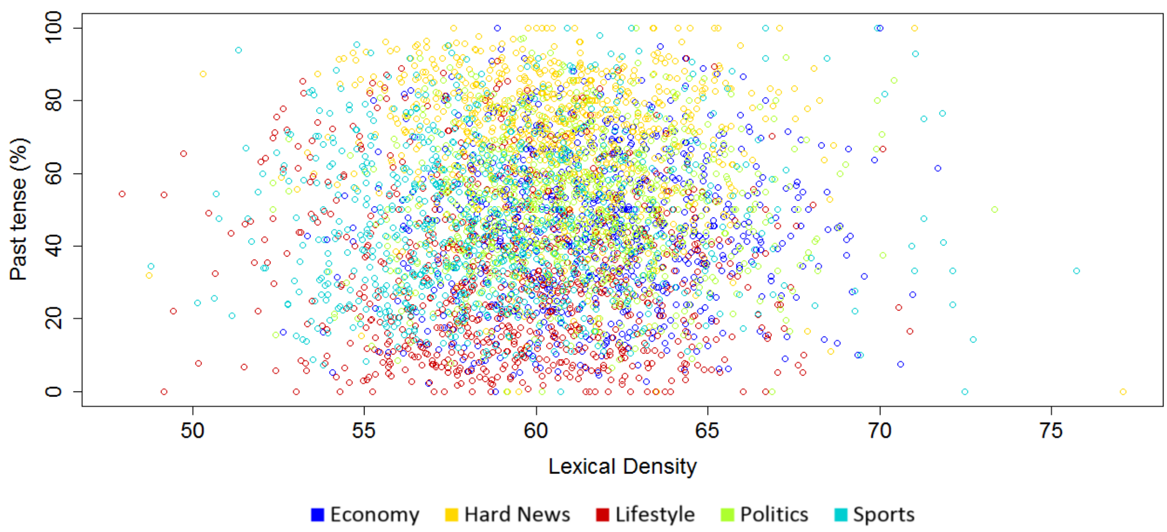


Figure 5.11: Lexical density versus past tense use

The distribution along the two dimensions shows that hard news articles form a cluster at high past tense values and a lexical density between 55% and 65%. The most centralised domain is politics, which contains hardly any texts outside its main cloud between 57% and 65% density and 40% to 70% past tense use. Economy is a little less focused, but produces variation mainly with regard to lexical density and towards the higher end. Sports articles fill the whole continuum and hardly depict any cluster, whereas lifestyle articles, although also widely distributed, clearly tend towards low past tense values and central to low lexical density.

In terms of goal orientation, or the degree of informativeness, this lower degree of standardisation of sports and lifestyle is confirmed in individual aspects; as the ranges for the respective linguistic features showed, sports news produces a particularly wide internal variation regarding the type-token-ratio, whereas lifestyle articles vary most in terms of declarative mood. Figure 5.12 visualises the differentiation of the individual domains, across all varieties.

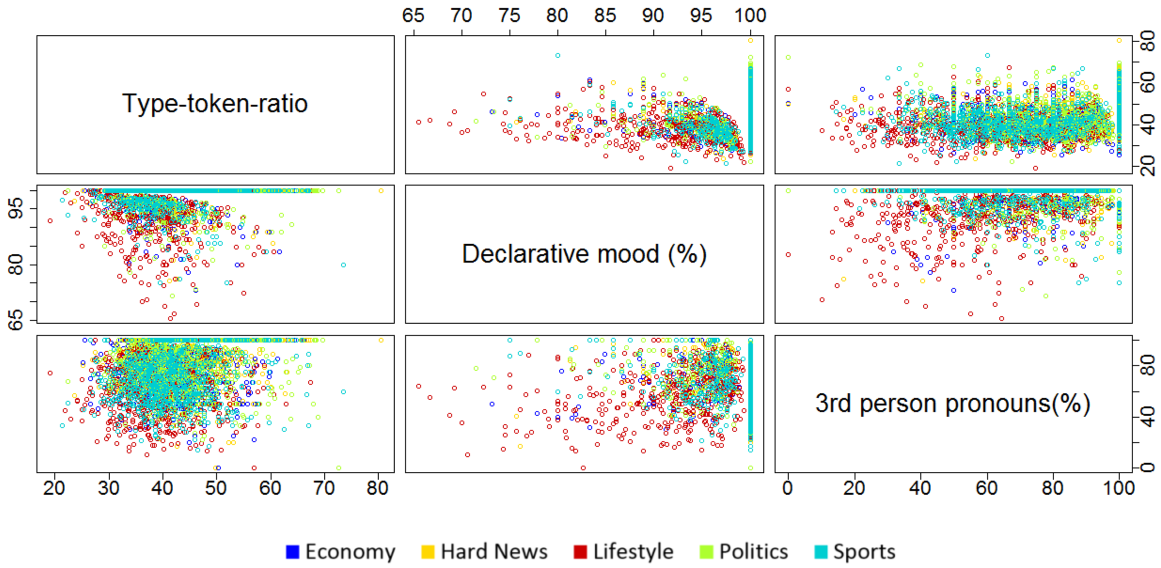


Figure 5.12: Correlations of indicators of goal orientation

The graphs in this figure represent the relationships between the three operationalisations of goal orientation, type-token-ratio, mood and 3rd person pronouns. The factor of mood is represented in this matrix by the percentage of declaratives, which worked as the operationalisation for informativeness. 3rd person pronouns are included with their percentages of all pronouns.

Since many categories, as was seen before, display hardly any imperatives or interrogatives, most texts in this dimension are clustered at the top around 100%. Among those that differ from this, sports articles form a small cluster around 95% and between a type-token-ratio of 30 and 45 and 50% to 90% 3rd person pronouns. The domain of lifestyle articles once again displays the broadest variance, reaching both high and low values in all dimensions. Although the other domains produce a number of texts each that disperse across the linguistic features, these instances are fewer, especially below 95% declarative mood.

For the degree of informativeness, all three features therefore indicate similar tendencies: lifestyle articles are least focused on the function of informing, sports news varies a little from the norm but are still more centralised than the former. The other three domains are less dispersed and the majority of texts can be found on the higher ends of the respective scales. Among the three linguistic features, both the use of 3rd person pronouns and declarative mood produce texts with maximum values of 100%, but also display lower values. Type-token-ratio naturally does not reach any maximum, but contains the bulk of texts in an interval between 35% and 50% and therefore offers less variation from which to draw conclusions.

5.2 Tenor of Discourse

The second parameter to define a situational context is that of the interpersonal relations between the participants. As was described in section 3.2.2.1 the subdimension of agentive roles is not relevant for this study; instead, the tenor will be analysed in terms of the social role relationship and the social distance.

5.2.1 Social Role Relationship

The social role relationship reflects the power structures that are present between the participants of a discourse situation (Steiner 2004, 17; Neumann 2013, 63). For this investigation, it has been operationalised to be represented by mood, modality, the amounts of keywords and mean word length. These linguistic features contribute to the definition of this parameter in different ways; modality and mood, especially in the form of interrogatives and imperatives, are reflections of differences in authority between the participants, while mean word length and the amount of keywords used in a domain are more general indicators for the use of technical language, which signals expertise on the side of the author as well as requires expertise from the reader.

Mood

The linguistic feature of mood was already part of the subdimension of goal orientation. In contrast to its function there, where the declarative mood was of major interest, at this point interrogatives and imperatives are the focus, as they reflect the interpersonal factor of the discourse.

In general, as became apparent in the previous section, these two types of mood are relatively rare in the texts analysed here. Of the 4,000 articles, 2,979 (74.48%) contain neither interrogatives nor imperatives. Of the remaining 1,021 texts, 119 are from the domain of economy, 105 from hard news, 152 from politics, 205 from sports and 440 from lifestyle.

Of the two types, imperatives are the ones with the clearer function, but also fewer instances in the corpus. The use of imperatives, as was discussed in section 3.2.2.2, is a strong indicator of authority, however, it functions differently in the various domains. In general, articles from the lifestyle domain show a significantly higher amount of imperatives than the other sections, as can be seen in Table 5.6.

Table 5.6: Frequencies of imperatives

	AUS	HK	KEY	UK	USA	∑ *% of domain*	*range*
Economy	17	10	7	28	29	91	22
	0.56%	0.29%	0.25%	0.68%	0.73%	0.53%	0.48
Hard News	21	3	20	12	12	68	18
	0.81%	0.1%	0.74%	0.28%	0.28%	0.4%	0.71
Lifestyle	195	82	125	100	135	637	113
	3.67%	1.37%	1.95%	1.58%	2.11%	2.1%	2.3
Politics	19	3	14	17	16	69	16
	0.59%	0.09%	0.45%	0.41%	0.38%	0.38%	0.5
Sports	16	15	23	23	39	116	24
	0.43%	0.4%	0.78%	0.45%	0.75%	0.56%	0.38
∑	268	113	189	180	231	981	
% of variety	1.5%	0.58%	1.05%	0.75%	0.96%	0.95%	
range	179	79	118	88	123		192
	3.24	1.28	1.7	1.3	1.82		3.58

Regarding the overall frequencies of the domains, the variation in terms of imperatives is highly significant at χ^2=604.71 (df=4, p<0.000), to which the frequent use in lifestyle articles contributes 419.46. This is reflected in the individual varieties as well, where this domain is clearly set apart from the others in each case. The other four domains behave differently depending on the region; economy articles show by far the lowest amount of imperatives in Kenya, whereas this form of mood is comparatively frequent in the UK and the USA. Hard news on the other hand is outstanding in Kenya and Australia, but contain hardly any imperatives in the other varieties.

When looking at the instances more closely, it becomes apparent that imperatives are used to different ends in the individual domains. The first obvious difference is the inclusion of quotations. The percentage of imperatives that are at the beginning of a direct quote are much higher in politics (29%, 20 instances), sports (31.03%, 36) and hard news (32.35%, 22) than in economy (19.8%, 18) and lifestyle (13.19%, 84). In these instances, the reflected authority is not the author's but the actor's. The respective instances show varying functions as well: in sports news, an imperative, positioned at the beginning of a quote, is often used to introduce a statement and weaken it, or ask for understanding or consent, as in the following example:

> [3] […] While Pulver rejected suggestions Wallabies players would be more motivated to perform if their finances were to suffer through a defeat, he did admit he's a fan of an incentive-based payment structure. "**Look**, in the ideal organisation - forget rugby - in any organisation I like the concept of performance-based pay," Pulver said. (Tucker 2013, *The Courier Mail*; emphasis added)

This usage in direct quotes cannot be found frequently in hard news, politics or economy, however, since the overall frequencies of imperatives are so low in these domains no conclusions can be drawn on this basis. A certain similarity can be observed regarding the verbs that are most frequent as imperatives: in news from politics, economy and sports, *let* occurs most frequently, in hard news, *look* is dominant, and in lifestyle, *try* and *get* are used most often. Here, too, the frequencies are very low however, so that these verbs can be regarded as no more than rough hints.

Among the varieties, the differences are again significant at a χ^2-value of 98.02 (df=4, p<0.000). Hong Kong contains the lowest frequencies of imperatives at χ^2=28.08; more significant is Australia's position at the other end of the scale, which shows a χ^2-value of 57.07. Especially in the lifestyle domain, the region's dominance can be traced back to a large degree to five texts which together contain 54 imperatives and therefore make up for 27.69% of all instances in this Australian domain. These relatively high frequencies are due to the structures of the texts; although dealing with different topics, they all include tips or advice, partly even in forms of lists, as in the following example:

[4] Solo female travel tips
Understand the cultural norms
Read about your upcoming destination, **email** local expats or locals who blog; **figure out** the geopolitics and religions and these will inform your travels as well as your behaviours. [...]
Involve others in your safety
Find ways to involve the people in this new place in your safety [...] **Tell** your hotel you're travelling alone [...]
Carry a doorstop and safety whistle [...]
Stay aware [...]
Stay sober [...]
Carry travel insurance [...]
Pay for your safety
Take a cab when lost or unsure where to go. **Spring** for the closer hotel. **Plan** so you're not in risky areas after dark.
(O'Donnell 2013, *The Courier Mail*; emphasis added)

Topics like this require imperatives and are often responsible for clusters of them in individual texts. This becomes more apparent when looking at the distribution of imperatives across the texts, depicted in Figure 5.13.

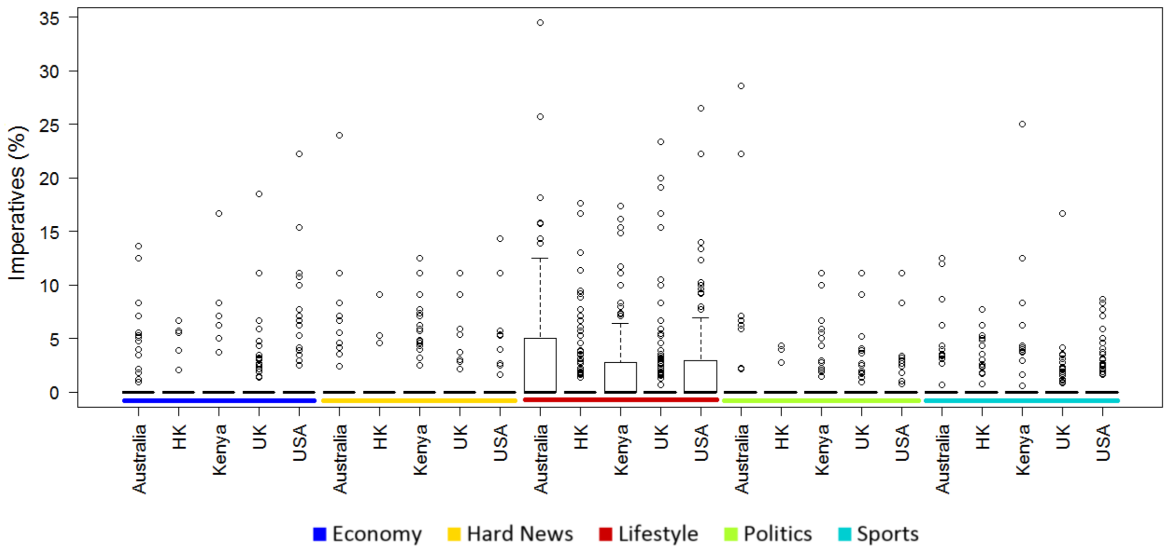


Figure 5.13: Distribution of imperatives

Due to the immense dominance of declarative mood, which was discussed in section 5.1.2, the median values of the categories are zero in all cases. Only lifestyle displays any even distribution beyond that in Australia, Kenya and the USA. In the other domains, the outliers represent those texts which include imperatives at all.

In the light of these results and the low overall frequencies, conclusions with regard to the exact function intended by imperatives are difficult. Lifestyle articles contain the highest numbers in all varieties, but specifically so in Australia, Kenya and the USA. In this domain, the imperatives are least often at the beginning of a quote, which means that they do, in fact, come from the author directly. Especially in this domain however, the advisory aspect of imperatives, as illustrated in the example above, is very strong, which makes them less the product of authority than of expertise, although the two dimensions of course contribute to each other. In this regard, the results match the comparatively low information value of lifestyle articles; the focus is not on purely passing on information, but also guidelines or directions. This requires an interpersonal dimension which is not necessary for an informative text, and also implies superiority of the author as well as its more explicit expression.

The case is different concerning interrogatives. In total, there are 1,645 question marks in the corpus, and although the dominance of the lifestyle domain is repeated here, more diversity can be found regarding the varieties, as Table 5.7 illustrates.

Table 5.7: Frequencies of interrogatives

	AUS	**HK**	**KEY**	**UK**	**USA**	∑ *% of domain*	*range*
Economy	18	8	13	78	28	145	70
	0.6%	0.23%	0.47%	1.9%	0.7%	0.84%	1.67
Hard News	7	13	8	45	36	109	37
	0.27%	0.44%	0.29%	1.05%	0.85%	0.65%	0.78
Lifestyle	150	97	163	290	218	918	193
	2.83%	1.62%	2.54%	4.59%	3.40%	3.02%	2.97
Politics	23	23	35	76	50	207	53
	0.71%	0.69%	1.12%	1.85%	1.2%	1.15%	1.16
Sports	33	29	26	97	81	266	71
	0.89%	0.77%	0.88%	1.88%	1.56%	1.28%	1.11
∑	231	170	245	586	413	1,645	
% of variety	1.29%	0.87%	1.36%	2.44%	1.72%	1.59%	
range	143	89	155	245	190		283
	2.56	1.39	2.25	3.54	2.7		4.36

Again, Hong Kong produces the lowest frequencies, but here, the UK is the variety with most instances. Almost half of the questions in this part originate from lifestyle articles, whereas hard news displays only few interrogatives. Kenya and Australia are very close together and differ mainly in articles dealing with politics, while the USA, especially with regard to the relatively high values in hard news and sports, is closer to the UK on the one hand and to Australia and Kenya with regard to the other domains on the other hand.

The dominance of the UK among the varieties and lifestyle among the domains is not due to chance. The variation between the regions is significant at χ^2=194.49 (df=4, p<0.000), to which the high values in the UK contribute 109.36. The differences between the domains are even more expressed at χ^2=587.83 (df=4, p<0.000), a value which can be traced back largely to the frequent use of interrogatives in lifestyle articles (388.04); the questions from the UK part make up for 31.6% of all in this domain. A qualitative approach to the texts in this category shows that 20 hits occur in headlines; these can be found to work as incentives to read the rest of the article as they summarise the topic and increase curiosity, as in examples [5] and [6], or to realise puns, as in examples [7] and [8]:

[5] Aaargh! The world's silliest haircut is back and it's huge on the catwalk. But can a real woman pull off a MULLET?
(Pullman 2013, *Daily Mail*)

[6] Bitcoins explained: A viable future currency, or an accident waiting to happen?
(Meade 2013, *Daily Mail*)

[7] What's got interim? Neuropsychologist analyses short-term Chelsea boss Rafa Benitez
(Weeks 2013, *The Sun*)

[8] The sexual politics of hair – are all cuts created equally?
(Adewunmi 2013, *The Guardian*)

These questions do not function as markers of authority, but rather as teasers for what is to follow in the article. Within the texts, questions sometimes display a similar function when positioned at the end of a paragraph to lead over to the next, and in many cases, the questions are answered right away. Sometimes, as the following excerpt shows, they are phased more personally and thus take the form of a partly rhetorical question:

[9] **For instance, how many people do you know still use an alarm clock?** Most people now tend to rely upon their phones, iPods or tablet devices to wake them up. (Barnett 2012, *The Daily Telegraph*; emphasis added)

In these cases, the interrogative does not require an actual answer, neither from the reader nor from the text itself, but is an alternative way to phrasing the content as a declarative and a means to make the text more geared towards the audience. These different ways of using interrogatives can be found in the lifestyle domains from the other regions as well, with the exception of puns, which hardly ever occur outside the UK. In this variety however, it occurs in all domains, preferably in headlines.

Apart from this British particularity, the domains and varieties show little variation beyond the mere frequencies; interrogatives are used either as more audience-oriented substitutes for declaratives or as teasers for the article itself, and often as a combination of both. Especially in headlines, they open up the topic to the reader and summarise the major issue, thus stimulating curiosity. In none of these functions however are interrogatives used as indirect imperatives or signal a difference in authority; rather, particularly rhetorical questions connect the discourse participants by showing that they share attitudes and experiences. With regard to the social role relationship they therefore create equality, but this way also contribute to the subdimensions of social distance and experiential domain as they construct closeness and relate to similarities.

Modality

The linguistic feature of modality was defined as the second operationalisation of social role relationship. In this context, the use of modals is seen to represent the weakening of a statement or signal insecurity about a topic or information. A high frequency of modals thus decreases the levels of authority and expertise constructed in the article.

In the tagset applied to the data at hand, all modal verbs are subsumed under the part of speech-tag MD. It includes the lemmata *can*, *could*, *do*, *may*, *might*, *must*, *need*, *ought*, *shall*, *should*, *will*, *would*, as well as the word parts *ca* and *wo* for the negatives *can't* and *won't* (the contraction *n't* is assigned a separate part of speech-tag throughout the whole dataset). In the calculations done here, these two negative forms were added to their respective positives. An overview as given in Figure 5.14 immediately shows that these verbs differ in their frequencies; as *do*, *shall*, *need* and *ought* in their modal forms only make up 0.006% of the corpus, they were summarised to be visible in the graph.

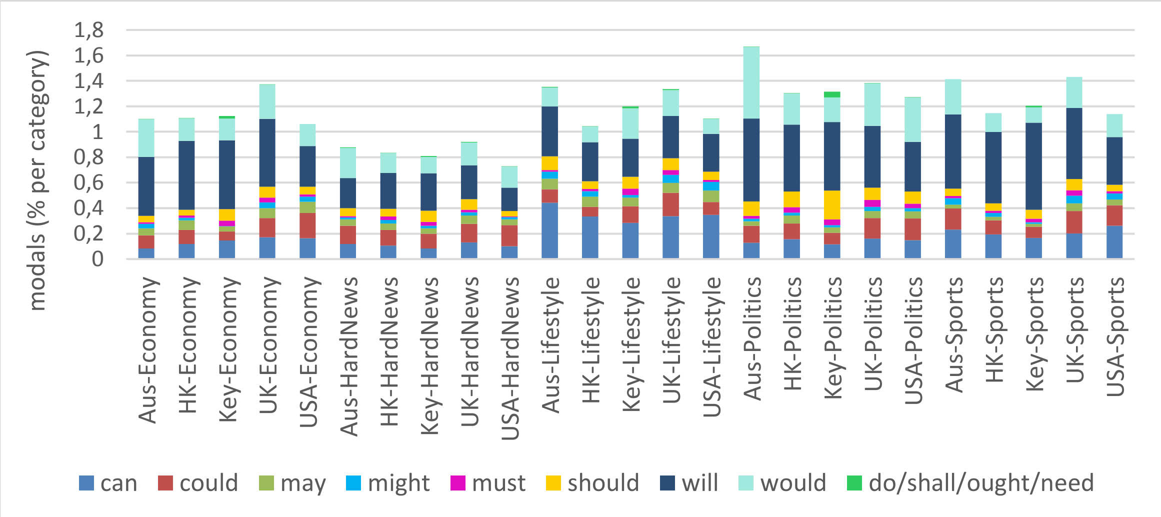


Figure 5.14: Relative frequencies of modals

Looking at these distributions, a number of particularities stand out right away: among the domains and across all varieties, hard news contains fewest modals, which confirms the previous results that these articles are focused on facts and therefore do not require the relativisation of statements as much as news from the other domains. Speculations are left out and articles are restricted to what is known for sure, which is also reflected in the fact that hard news articles are on average the shortest. The differences between the domains in terms of modality are significant at χ^2=538.83 (df=4, p<0.000), and the low use of modals in hard news is mirrored in a χ^2-value of 359.57.

Among the other domains, economy displays a very constant picture with the exception of news from the UK. Lifestyle and sports are more diverse internally, although it can be seen in both domains that Australia and the UK are very similar in producing high values, while the other varieties also form a cluster around lower values. The domain with the highest overall frequencies of modals is politics, which exceeds what was statistically expected at χ^2=141.13. The dominance of politics in this aspect indicates that especially political topics are often kept vague or include promises rather than facts. The modals most frequently used are *will, would*, *can*, *could*, but also *should*, which is less relevant in the other domains and underlines the assumption of increased vagueness in this domain as it is less concrete and committal than for instance *will* or *can*.

While *will*, *would*, *could* and *can* are the strongest modals in all categories, the latter stands out particularly in lifestyle articles. *May* and *might* also have higher frequencies here, but the differences are not significant. The varieties vary less than the domains but are still significant at χ^2=196.39 (df=4, p<0.000), with Australia containing most modals (1.31%) followed by the UK (1.3%), Kenya (1.15%), Hong Kong (1.09%) and the USA (1.07%). The discrepancy which sets Australia apart is most prominent in the domains of lifestyle and politics, as a look at the distribution of modals across all texts shows (see Figure 5.15).

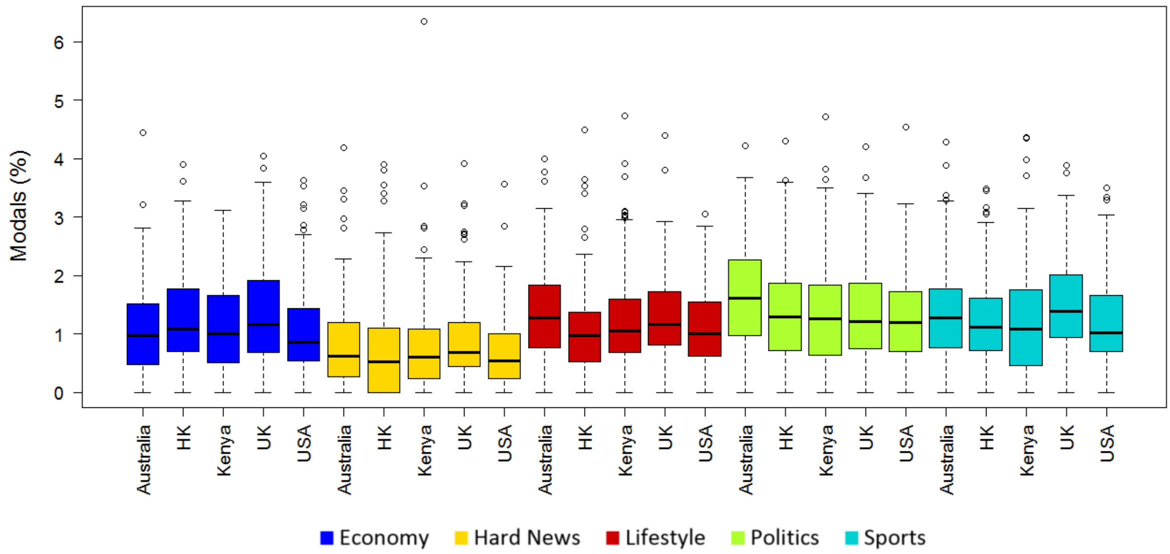


Figure 5.15: Distribution of modals

In every category, there are texts which do not contain modals at all, and since they are not outliers but form the boundary of the lower whisker (or even of the main box in the case of hard news from Hong Kong), it can be assumed that using no modals is not an irregularity in news. The most constant values occur in the domain which uses them most however, namely in political news. The medians of all varieties except Australia are almost identical. In hard news, all five median values are very close, yet the varieties vary more in this domain regarding their overall distribution. Lifestyle, economy and sports show most internal variation, whereas among the varieties, Australia stands out in this respect.

With regard to the social role relationship for which modals are an indicator here, these results can be summarised first of all by saying that again, there is more variance between the domains than between the varieties. Especially in political news, it is doubtful whether modals are used to cover insecurities about a topic; it seems more likely that here, their main function is to weaken the statements made by political actors and avoid commitment. Since the corpus is not annotated for quotations, it cannot be queried how many instances of modal verbs occur inside direct quotations; nevertheless, including the modal can still be regarded as a choice made by the author instead of rephrasing the respective statement in a more concrete manner. In general, modality can thus be seen as a conscious strategy to avoid constructing too much authority, which would automatically make a statement more substantial and relevant.

The opposite of this constructed vagueness can be observed in hard news, which contributes the lowest numbers of modal verbs in all varieties. As was already said above, this can be traced back to the focus of this domain on facts and the low degree of speculation. The author needs to retain both authority and expertise to count as a valid source of information, and modality would undermine this image.

In contrast to the domains, the regions are more evenly distributed. What can be said is that the tendency shown for interrogatives and imperatives, especially regarding the low use of these features in Hong Kong news, is not confirmed here. While interrogatives were analysed as markers of a more equal social role relationship, rendering Hong Kong the most unequal variety in these terms, the region does not display particularly few modal verbs to confirm this trend. This is even truer for the USA, which contains relatively many interrogatives but the fewest modals of all regions.

Amount of keywords

In contrast to the factors of mood and modality, the last two operationalisations are more general measures which will complement the former two in an overall analysis of social role relationship. One of these is the amount of keywords. The exact lists of which words are characteristic for the individual categories were discussed in section 5.1.1 (see also appendix 3), but apart from a qualitative analysis on the lexical level, the mere numbers of such significant words are an indicator for the specialisation of the vocabulary. If a word produces a high keyness value for a certain domain or variety, that implies that it occurs significantly more often in this category than elsewhere and therefore constitutes an instance of language for specific purposes.

As was said before, the critical value for the threshold of what counts as a keyword was defined as 15.13 ($p < 0.0001$). This is relatively high, and many studies work with 3.84 ($p < 0.05$) or 6.63 ($p < 0.01$) for this purpose. For the present study, the higher value was preferred as the interval between 6.63 and 15.13 was responsible for more keywords in all categories than the interval ≥15.13, despite the latter covering a much larger range in terms of keyness values. This would have skewed the calculations towards the lower end, and since especially high keyness values reflect specialisation, the higher threshold was applied. For all categories, the keywords were counted both in variety-internal and domain-internal comparisons, based on the lists in appendix 3. Table 5.8 summarises the results for the domain-internal comparisons, listing the number of keywords, the percentages of all word types that are key, their cumulative frequencies and the percentages of these in relation to all word tokens of the category.

Table 5.8: Amount of keywords - Domain-internal dimension

Economy					
	AUS	HK	KEY	UK	USA
No. of keywords	239	328	361	224	287
% of all word types	3.13	4.15	4.96	2.57	3.29
Cum. frequencies of keywords	8,623	14,334	20,131	9,705	12,853
% of all word tokens	12.13	17.49	30.64	11.48	15.08
Hard News					
	AUS	HK	KEY	UK	USA
No. of keywords	249	351	367	260	339
% of all word types	3.55	4.34	5.04	2.88	3.75
Cum. frequencies of keywords	7,578	10,553	15,892	12,809	12,403
% of all word tokens	13.83	16.25	27.23	15.51	14.54
Lifestyle					
	AUS	HK	KEY	UK	USA
No. of keywords	246	393	468	249	357
% of all word types	2.05	2.82	3.54	1.90	2.47
Cum. frequencies of keywords	11,774	19,561	33,606	13,951	13,211
% of all word tokens	11.62	16.63	26.58	12.22	11.23
Politics					
	AUS	HK	KEY	UK	USA
No. of keywords	306	360	482	230	310
% of all word types	3.86	4.19	6.11	2.43	3.22
Cum. frequencies of keywords	14,680	14,691	22,291	12,955	13,217
% of all word tokens	19.16	17.85	29.73	14.63	14.84
Sports					
	AUS	HK	KEY	UK	USA
No. of keywords	316	345	591	224	361
% of all word types	3.84	3.79	7.35	2.28	3.96
Cum. frequencies of keywords	11,650	21,552	19,312	14,851	17,883
% of all word tokens	14.53	24.95	29.83	14.87	18.30

This overview allows a number of conclusions; the first obvious result is that in all domains, Kenyan news produces the highest values for keywords both in absolute and relative terms. The difference to the other varieties is particularly outstanding in political and sports news, where 6.11% and 7.35% of all word types in the respective Kenyan parts are keywords. Concerning the percentages these keywords contribute to the whole word count, economy, hard news and lifestyle items display the dominance of Kenya as well. These extreme values can be traced back to place names; as was seen in section 5.1.1, Kenyan news, like all other news, refers to prominent places in the country fairly often, and apart from this also include mentioning of countries and cities nearby. In the case of Kenya, these names hold high keyness values not necessarily because they are used more often than respective place names in the other regions, but because these particular places, namely *Kenya*, *Nairobi*, *Uganda*, *Africa* etc., hardly occur elsewhere. Names like *USA*, *UK, America*, *New York* or *London* on the other hand are common throughout all regions, and therefore do not result in such high keyness values.

Although this has an influence on Kenya being so outstanding in this regard, it should be noted that place names, too, imply specialisation. *Kenya* and *Nairobi* for instance are not technical terms as such and are most likely familiar to most people, yet details about these places are less commonly known and understanding any information given on them in a news item requires a certain minimum expertise from the reader. This becomes more expressed the more local the place references become, while especially in the USA and Australia, several bigger cities are prominent worldwide and feature in news internationally.

Another aspect that is observable from the table is that in lifestyle articles, the percentages of word types that are key are generally lowest and the differences between the varieties smallest, which in terms of the domain-internal comparison means that the news items from the different regions are most similar in this domain. The largest differences are to be found within sports news; these articles are thus very specialised on their respective region, which is not surprising seeing that different sports are popular in these societies and the national leagues are usually dominant. A high degree of specialisation is also displayed in political news, where of course internal political matters and national politicians and parties are key. Looking at the distribution of keywords from the other angle, namely in variety-internal relation, many of these results are mirrored, as Table 5.9 shows.

Table 5.9: Amount of keywords - Variety-internal dimension

Australia					
	ECO	HN	LIFE	POL	SPO
No. of keywords	481	508	462	364	566
% of all word types	6.29	7.24	3.86	4.59	6.89
Cum. frequencies of keywords	22,970	16,372	26,963	24,525	22,552
% of all word tokens	32.32	29.89	26.60	32.01	28.13
Hong Kong					
	ECO	HN	LIFE	POL	SPO
No. of keywords	508	470	581	416	463
% of all word types	6.43	5.81	4.17	4.85	5.08
Cum. frequencies of keywords	26,046	13,478	29,530	22,912	25,730
% of all word tokens	31.78	20.75	25.11	27.84	29.79

Kenya					
	ECO	HN	LIFE	POL	SPO
No. of keywords	513	437	433	415	547
% of all word types	7.04	6.00	3.27	5.26	6.80
Cum. frequencies of keywords	20,391	19,472	36,358	24,919	18,001
% of all word tokens	31.04	33.37	28.76	33.23	27.81
UK					
	ECO	HN	LIFE	POL	SPO
No. of keywords	515	515	432	408	573
% of all word types	5.91	5.70	3.30	4.31	5.84
Cum. frequencies of keywords	29,197	24,380	27,367	24,211	27,087
% of all word tokens	34.53	29.52	23.98	27.34	27.12
USA					
	ECO	HN	LIFE	POL	SPO
No. of keywords	618	649	556	474	667
% of all word types	7.09	7.19	3.85	4.93	7.32
Cum. frequencies of keywords	26,340	25,420	29,896	22,442	37,803
% of all word tokens	30.90	29.80	25.42	25.20	38.69

The figures show that within every variety, lifestyle articles produce the lowest numbers of keywords, although in Hong Kong, their cumulative frequencies make up a larger portion of the category in total than in hard news. The highest average of keywords across the domains can be found in the USA with 6.08%, in contrast to the UK which shows the lowest average at 5.01%. The specialisation of the articles therefore seems to be relatively high within the USA and the variation between the domains is significant at χ^2=210.06 (df=4, p<0.000), while the domains in the UK are less characteristic for their individual topics and share more general content. Nevertheless, even here the variation shows a χ^2-value of 129.32 (df=4, p<0.000), making it strongly significant still. The most extreme range between the numbers of keywords occurs in Kenya, with a difference of 3.77% between the domains of economy and lifestyle; economy, hard news and sports are particularly specialised here, while lifestyle again appears more general.

The differences between the domains are significant in every variety, but particularly the trend towards low numbers of keywords in the domain of lifestyle, which already showed in the previous comparison, shows high significances with χ^2-values ranging from 24.52 in Hong Kong to 107.92 in Kenya. It can thus be safely regarded as containing the lowest degree of specialisation, which confirms the results for the previous linguistic features. With a focus on interpersonal matters rather than on facts and content, a specialisation is less relevant and the articles are more accessible for an audience without any particular expertise. More keywords than statistically expected are demonstrated by the domains of economy, hard news and sports, the latter with the exception of Hong Kong, where this domain does not show any particular specialisation in relation to the others.

In contrast to the other operationalisations in this subdimension, numbers of keywords were not calculated per text as there is no straightforward way to approach this; because of the different directions of the comparisons, every category results in two lists of keywords, one for the variety-internal and another for the domain-internal perspective, and both lists would have to be queried separately in every text in order to allow valid conclusions in the end, an endeavour which would have gone beyond the scope of this work. The generating of keyword lists for every individual text on the other hand would have given rise to the question of what to relate the text to – the category, the domain, or the variety. Every one of these comparisons would have been useful in its way, but none would have given a complete picture. Relating a text to all other 3,999 articles however would not be very useful, as every text is of course specialised on the topic it treats and the amount of keywords and the keyness values they display would be very dependent on the text length. The feature of keyword numbers thus has to be regarded on the level of domains and varieties for this investigation.

Mean word length

The last operationalisation for the social role relationship is the mean word length. This measurement serves as an indicator for specialised language in the sense that technical terms on average are longer than less specialised ones (Bartsch 2009, 115). This, of course, is a generalisation and the degree of specialisation should not be interpreted solely on the basis of this value; yet, in combination with the previous linguistic features it is a valuable completion of this subdimension and, as the last feature in line, can be put into context.

The mean word length was calculated as the number of characters per word, not including punctuation or spaces. Symbols (part of speech-tag SYM), too, were excluded, as the vast majority are square brackets (1,279 of 1,359, 94.11%), slashes (31, 2.3%) and plus-signs (16, 1.18%). Table 5.10 shows the average mean word length values per domain and variety.

Table 5.10: Mean word length

	AUS	HK	KEY	UK	USA	*domain*	*range*
Economy	4.85	5.02	5.05	4.77	4.83	4.9	0.28
Hard News	4.71	4.9	4.87	4.64	4.82	4.78	0.26
Lifestyle	4.64	4.8	4.67	4.55	4.71	4.67	0.25
Politics	4.85	5.07	4.95	4.77	4.91	4.91	0.3
Sports	4.51	4.66	4.74	4.44	4.46	4.55	0.3
variety	4.7	4.88	4.83	4.62	4.74	4.75	
range	0.34	0.41	0.38	0.33	0.45		0.63

This overview over the categories illustrates that the differences in terms of mean word length both between the varieties and the domains are very small. Sports displays the lowest average word length among the domains, while economy and politics display the highest. Among the varieties, the UK produces the lowest, Hong Kong the highest value. These values are fairly crude as they are based on values for individual texts that are already averages of their respective words, and an analysis of every individual text in terms of its range is not possible here; nevertheless, a look at the distribution of the text-based averages gives more insight into the matter.

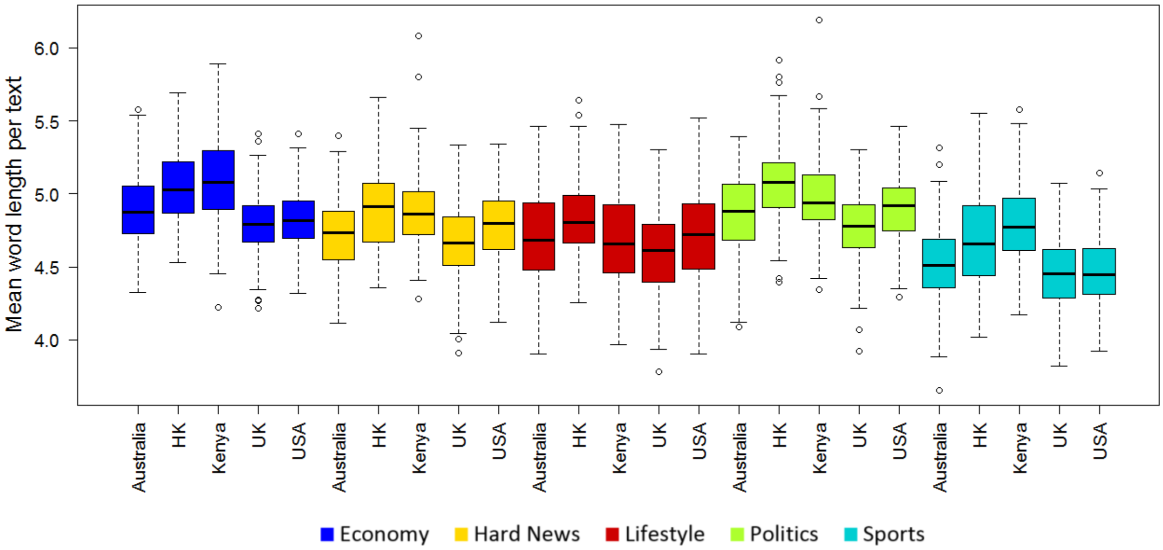


Figure 5.16: Distribution of the mean word length

The medians of the categories are very close to the averages displayed in Table 5.10, however, the domains and varieties show different ranges. The table gives a first indication of the internal variation of the domains and varieties, yet the boxplot representation depicts more clearly that the overall ranges are subject to some extreme outliers. Without these and on the level of individual texts, especially in lifestyle articles the mean word length differs a lot in Australia (with a range of 1.56), the UK (1.52) and the USA (1.62), while economy and hard news display the most compact plots on average with ranges of 1.28 and 1.41 respectively. Economic and hard news thus seem to be most standardised, whereas lifestyle articles, as was seen before, are more liberal in their structure. Furthermore, their low average value in terms of word length confirms the results regarding the keyword numbers that this domain is the least specialised one. For economy and hard news on the other hand, the relatively long words are in line with their high numbers of keyword, signalling a stronger specialisation here.

Regarding differences between the varieties, the two L2 varieties, Hong Kong and Kenya, produce the highest values for word lengths both on average and within the domains, with the exception of lifestyle. In the case of Kenya this fits to the high numbers of keywords, while with regard to mood and modality this variety did not stand out much. For Hong Kong, the observations are vice versa – the news displayed the lowest numbers for interrogatives and imperatives and showed low usage of modal verbs as well, thereby indicating that a certain degree of authority is constructed and an interpersonal focus is rarer than elsewhere, which is compatible with a specialisation of vocabulary. At the other end of this scale are the news items from the UK, which included by far the most interrogatives and also featured frequent modals, results which make the low average word length in this variety less surprising.

5.2.2 Social Distance

Apart from insights into the relationship between the social roles of the participants, the distance or closeness created between them is of interest for the tenor of discourse. This is reflected in the subdimension of social distance. As became apparent already in the previous section, the two dimensions of the tenor are very closely linked and sometimes hard to distinguish, which is why the results from the analysis of the social role relationship will be taken up again here at some points.

It was defined in section 3.2.2.3 that social distance will be operationalised in terms of evaluative language, place references, contractions, voice, 1st and 2nd person pronouns and nouns and nominalisations. The first parameter, evaluative language, which is represented by the use of forms of address and titles, boosters and minimisers, modals and adjectives and adverbs, brings with it a higher degree of subjectivity as some of the items to be queried are not predetermined by a word class or part of speech-tag, but have to be defined manually for the analysis. The description of the results will start with this parameter, and all decisions will be described and explained.

Evaluative language –Titles and forms of address

The use of titles and different forms of address can have various effects in news. In general, the inclusion of such items can indicate respect or seriousness of topic, but they can also serve as explanations, as in the following example:

> [10] Meetings with Guy Verhofstadt, **prime minister of Belgium** and holder of the EU presidency, and Romano Prodi, **president of the European commission**, included sessions on energy and plans for a "common European economic space" to promote trade. (Black 2001, *The Guardian*; emphasis added)

In cases like the ones shown here, the terms *prime minister* and *president* are included primarily to tell the reader who the actors in this sentence are, which illustrates that this knowledge is not presumed in the audience. The importance of these actors is relevant in order to justify the news worthiness of the story and the validity of the people as sources. Such descriptions of actors are not restricted to political news of course, but count as general practice of journalism (Harcup 2005, 109; Kolodzy 2006, 66). Although instances like this therefore do not serve as a title per se but are given for other reasons, they were included in this analysis as they still put an emphasis on the function and position of the respective people. Especially *president* is generally capitalised when used in front of a name as a title, and can thus be told apart from explanatory instances of the word. Altogether, the investigation of titles and forms of address includes the following search terms:

General:	*Mr, Mrs, Ms, Miss*
Academic:	*Dr, Doctor, Prof, Professor*
Political:	*Minister, PM, Chancellor, President*
Royalty / Nobility:	*Lord, Sir, Prince, King, Queen*

In the cases of *minister*, *chancellor* and *president*, the terms were searched only in their singular forms as the few instances of the respective plurals did not apply to concrete people, but were more general descriptions. The terms of the last category were sorted manually, as they included various hits of addresses (e.g. “King George Square”, “Prince Edward Road”), buildings (e.g. “Queen Elizabeth Hospital”, “King’s College”) or names like “Burger King” or the band “Queen”. The items *mister*, *Pres.*, *Madam(e)* and *Princess* were also queried, but produced no relevant hits.

The categories suggested in this list are not to be taken as conclusive, but merely serve to give a better overview. In total, the cumulative hits of all these items make up less than 1% of the whole corpus, which means that when expressed in percentages, the differences between the individual lemmata are very small. Summarised in categories, they become more easily traceable. Figure 5.17 illustrates the distribution across the different categories.

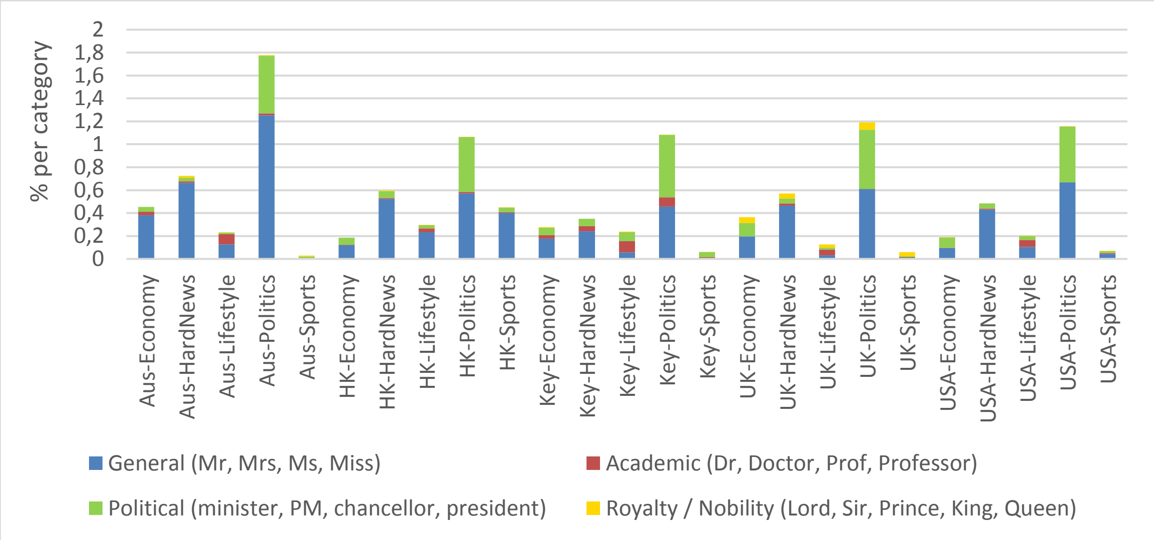


Figure 5.17: Relative frequencies of titles and forms of address

As can be seen immediately, the use of titles and personal addresses is not constant across the domains and varieties. In every region, political news produces by far the highest numbers, followed, also in every region, by hard news. In all domains except politics, the focus is on the general addresses of *Mr*, *Ms*, *Mrs* and *Miss*, of which the first, *Mr*, makes up 80.7% on average. The dominance of the male form here is particularly expressed in sports news and economy with averages of 89.47% and 89.87% respectively and is more moderate in lifestyle articles with an average of 67.2%.

In political news, *minister*, *president*, *PM* and *chancellor* are more relevant than elsewhere, which is of course predefined by the topic. Of the 1,345 instances of *president*, 618 are used as direct titles in combination with names, the most frequent ones of which are *Obama* (82), *Uhuru* (74), *Kenyatta* (66), *Barack* (41) and *William* (Ruto, 32). For Barack Obama the exclusion of the first name thus seems to be acceptable; independent of the item *president*, *Obama* appears 495 times in the whole corpus, only 62 of which are preceded by *Barack*. This might be due to the low ambiguity of the name; in comparison, *Bush* appears 108 times in total and 50 times in connection with a first name, of which there are seven different ones.

Other political figures which feature prominently in the data are treated differently; for *Putin*, there are 114 hits, the majority of which collocate with *Vladimir* (32) or *Mr* (38). Only five occur in connection with *president* or *ex-president*. A similar distribution can be found for *Sarkozy* (80 hits, 31 with *Mr* and 23 with *Nicolas*, 2 with *president*), *Hollande* (87 hits, 35 with *Mr* and 30 with *François*, 0 with *president*) and *Cameron* (278 hits, 93 with *David*, 88 with *Mr*, 1 with *minister*). The term *president* in English-language news thus seems to be connected primarily to the US-American president.

With regard to the varieties, it can be seen that especially Hong Kong news features both titles and addresses in all news domains. Sports is outstanding here in comparison to the other regions, where the values in this domain are close to 0. As was seen in the keyword lists in section 5.1.1, this particular result is a little misleading as the vast majority of hits for Hong Kong news in general come from the same newspaper, the *Wall Street Journal Asia* (1,946 of 2,147, 90.64%). In the USA, too, the *Wall Street Journal* contributes by far most titles and addresses (1,508 of 1,909, 78.99%), and in Kenya, the *Daily Nation* produces most instances (892 of 1,540, 57.92%). In the UK and Australia, the dominant papers are less explicitly set apart, with the *Telegraph* contributing 34.49% (695 of 2,015) of the British hits and the *West Australian* and *Sydney Morning Herald* each reaching 27.42% (640 of 2,334) in Australia.

Another observation which can be made on the basis of Figure 5.17 is that academic degrees feature most prominently in lifestyle articles. This trend holds true for every variety. The most frequent item in this category is *Dr*, which indicates that the relevance of this category for this domain might originate from articles dealing with health advice or beauty topics, and that the hits are therefore not evenly spread over the news items. Indeed, only 107 of the 800 lifestyle articles contain an instance of *Dr*, the top ten articles of which contain 31.2% of all hits. This is reflected in the overall distribution of titles and addresses on the basis of the individual texts, as depicted in Figure 5.18. This perspective clearly shows that across all varieties, lifestyle articles display a fairly small range of an even distribution, and that many articles are depicted as outliers. Furthermore, it can be seen that Australian news produces fewest outliers in all domains and instead shows the highest ranges, which reflects the even usage of titles and addresses in all newspapers.

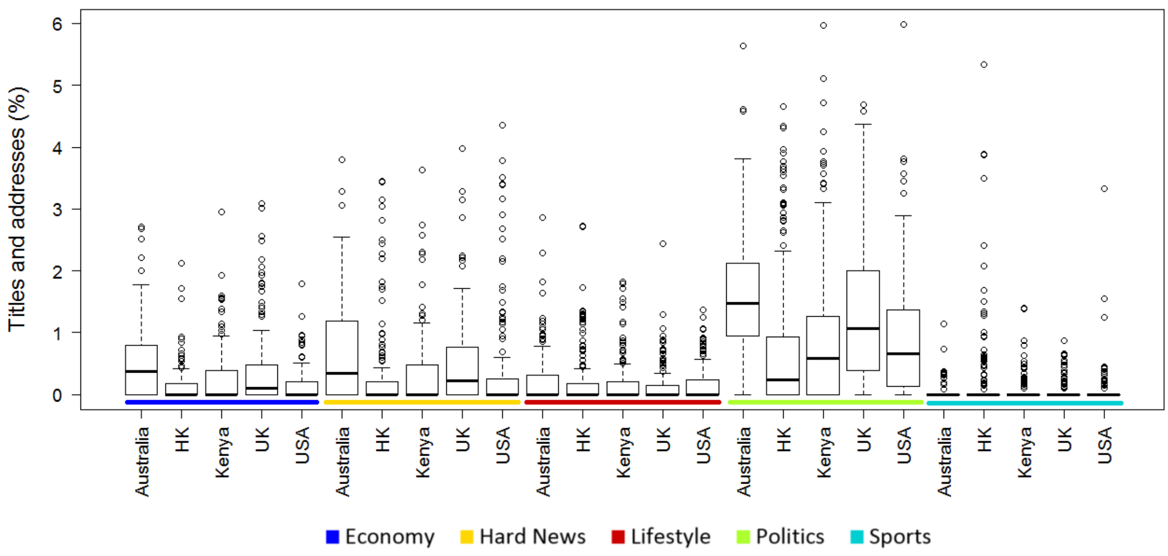


Figure 5.18: Distribution of titles and addresses

Regarding the larger framework of evaluative language, the results seem to indicate that in political news, there is a stronger tendency towards correctness. Actors are often referred to as *Mr*, *Mrs*, *Ms* or *Miss* X, rather than just using their last names. Concerning titles and political functions, the majority of these is of explanatory use, as the figures for *president* indicated. In hard news articles, the most dominant forms of address are the general ones which cannot be used as explanations that easily and thus point towards the construction of seriousness in this case. The names which collocate with *Mr*, *Mrs*, *Ms* and *Miss* in these news items are rarely those of prominent people or celebrities but the ones of those involved in crimes and accidents. A superficial scan of the hard news items of the corpus hints at a trend that especially victims are referred to with a formal address while convicts or suspects are not; reliable conclusions on this however would require in-depth analyses of the articles and can therefore not be drawn here.

In lifestyle articles on the other hand, the use of forms of address and especially titles seems to have a different function, namely that of establishing authority of the reported facts or pieces of advice and credibility of the source. This form of 'title-dropping' is common in popular texts dealing with more or less scientific topics (Conrad 1999, 291; Plesner 2012, 676), like articles concerning health or nutrition issues. All in all, titles and forms of address can thus not be assigned to one particular function, but can serve as markers of more neutral or correct style as in politics or hard news, or as emphasis of authority, which would be more relevant for the subdimension of social role relationship. On the other hand, leaving out these linguistic markers then results in a less neutral style and a less distant image, sometimes even degradation, of the respective people in the story. Despite the varying functions of these features, which should be kept in mind for the discussion, their inclusion can therefore be taken as a marker for decreasing evaluation in language and an increasing distance between the discourse participants.

Evaluative language – Boosters and minimisers

The aspect of boosters and minimisers has been included as an operationalisation for social distance as frequent use of these items has been found to indicate casual style and thereby lower the distance between the participants (cf. Eggins and Martin 1997, 232). They increase the degree of evaluation of a text as they modify the meaning of the boosted or minimised concept in the sense of the author. As was already said in section 3.2.2.3, numerous lists and suggestions can be found in the literature for lexical items which can function in this way. For the present investigation, the most frequently suggested boosters and minimisers each were included. Table 5.11 lists all items and their absolute frequencies in the whole corpus.

Table 5.11: Frequencies of boosters and minimisers

Boosters			**Minimisers**		
Lemma	Freq.	%	Lemma	Freq.	%
absolutely	75	0.003	*barely*	88	0.004
entirely	45	0.002	*fairly*	59	0.003
exceedingly	3	0.000	*hardly*	61	0.003
extremely	120	0.006	*just*	2,531	0.118
fully	153	0.007	*merely*	39	0.002
immensely	8	0.000	*nearly*	397	0.018
more	5,263	0.245	*only*	2,151	0.1
most	2,110	0.098	*pretty*	226	0.011
really	786	0.037	*quite*	259	0.012
thoroughly	24	0.001	*relatively*	137	0.006
totally	50	0.002	*scarcely*	5	0.000
utterly	10	0.000	*simply*	213	0.01
very	1,514	0.07	*slightly*	132	0.006
∑	**10,161**	**0.472%**	**∑**	**6,298**	**0.293%**

All calculations were done excluding hits in which the respective lemma was preceded by the *no, n't* or *not*, as the effects of the negation varied and the cases could therefore not simply be sorted into either minimising or boosting, as the following examples show:

[11] People believe that those who wear glasses look **not only** more intelligent but more honest. (Black 2011, *Daily News*; emphasis added)

[12] A Google search for "fake nerds" will convince you that, **not only** are they a problem, they're an epidemic. (Mouton 2013, *USA Today*; emphasis added)

In the first case, the *not only… but*-construction could be replaced by *and*, rendering the sentence as "People believe that those who wear glasses look more intelligent and more honest", as *honest* is not an increase of *intelligent*, but a different attribute. In example [12] however, *not only* introduces a climactic development, as an *epidemic* is a type of *problem* on a larger scale. While the boosting function in the first case is thus debateable, it is clear in the second case. In total, there were 657 hits of collocations with *no* or *not*, most frequently with *only* (219) and *just* (209). Furthermore, cases of *any more* were excluded as well, which occurred 56 times.

When looking at the distribution of minimisers and boosters across the domains and varieties, differences can be found for both dimensions (see Figure 5.19).

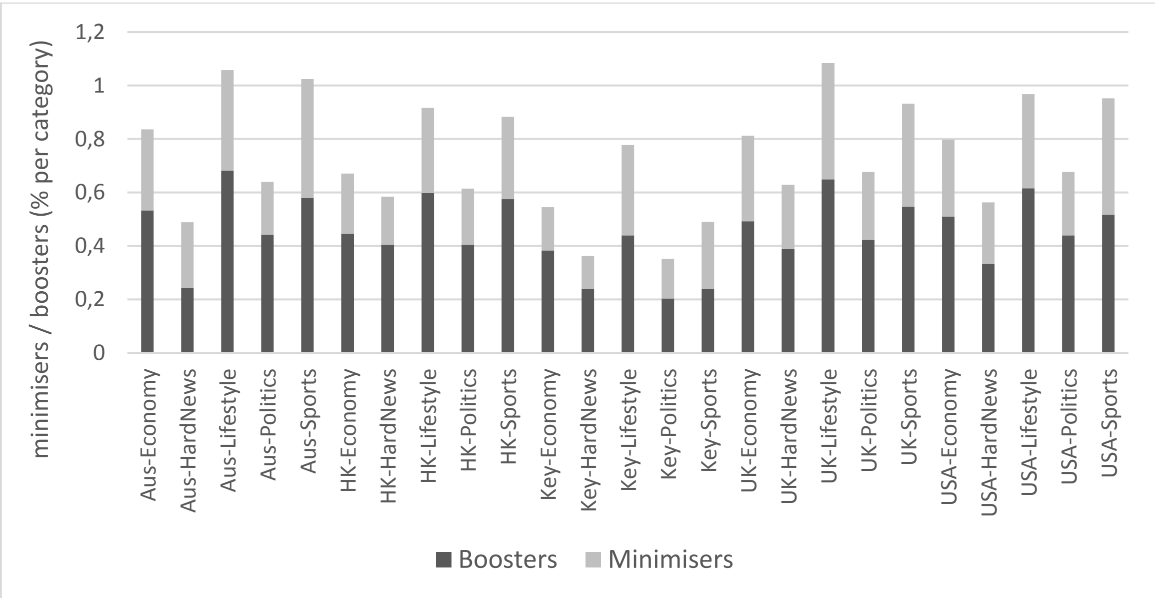


Figure 5.19: Relative frequencies of minimisers and boosters

In every variety, lifestyle articles contain the highest cumulative numbers of minimisers and boosters. In all regions except Kenya, they are followed by sports articles. Hard news and politics generally produce the lowest values, although the two domains are further apart from each other in Australia (0.15 percentage points difference) and the USA (0.11) than in Hong Kong (0.03), the UK (0.05) and Kenya (0.01). The differences between the domains are significant with χ^2=734.1 (df=4, p<0.000), which is mainly due to the high numbers in lifestyle (χ^2=271.1) and the low numbers in hard news (χ^2=234.11). Among the varieties, Kenya contains by far the fewest instances of minimisers and boosters with an average of 0.51%, followed by Hong Kong (0.73%), the USA (0.8%), Australia (0.81%) and the UK (0.83%). The variation between the varieties is highly significant with χ^2=328.82 (df=4, p<0.000), with the low values in Kenya alone reaching χ^2=242.24.

Regarding the differences between the two categories of minimisers and boosters, the distribution across the domains and varieties reflects what was already indicated by the absolute numbers in Table 5.11, namely that boosters are by far more common. The only exceptions to this are hard news in Australia and sports news from Kenya, where minimisers are very slightly more frequent (0.004% and 0.01% difference respectively). These two phenomena, and the low numbers of Kenya in general, are not exceptions within their categories, as the distributions across all texts in Figure 5.20 illustrates.

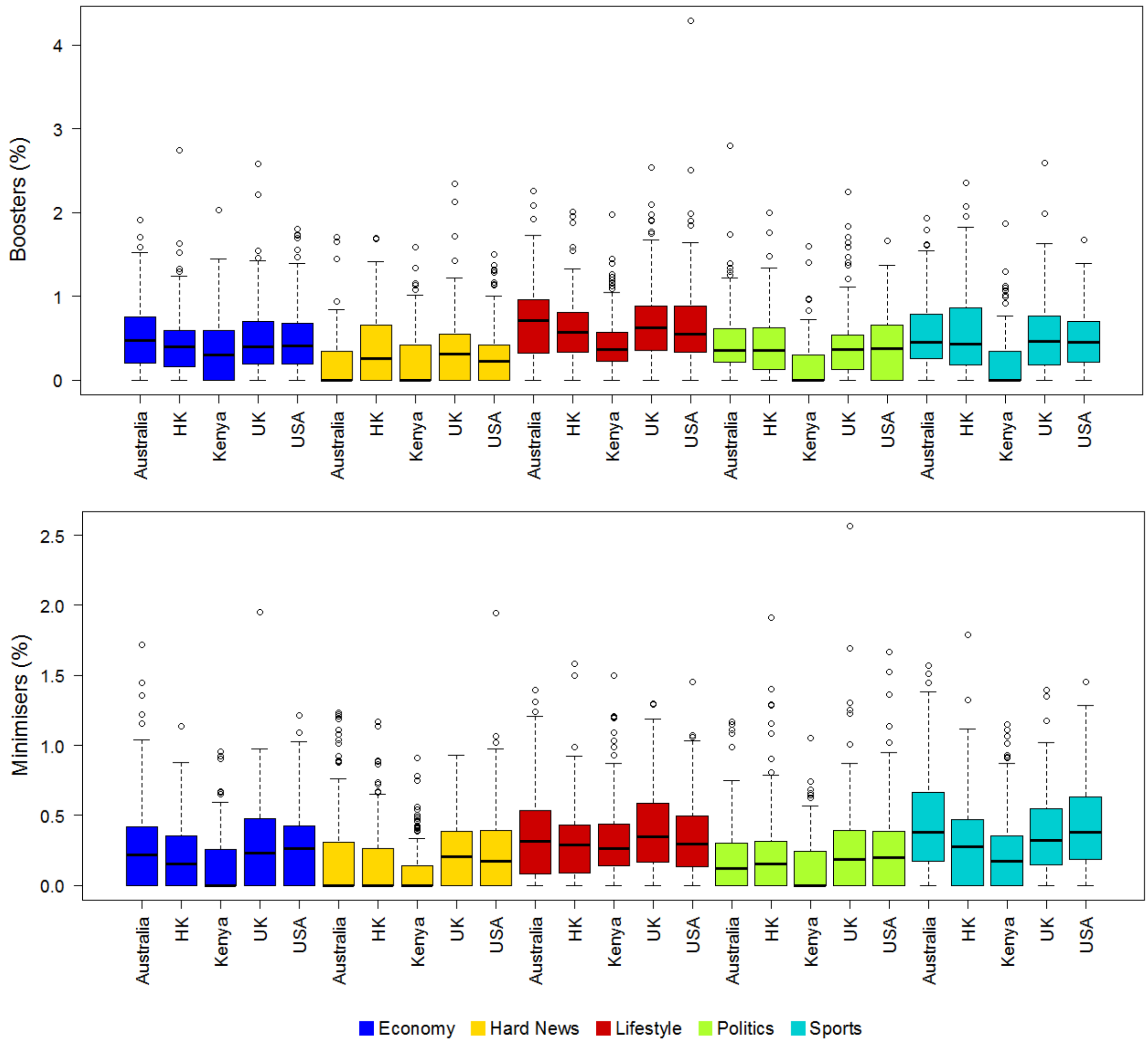


Figure 5.20: Distributions of boosters and minimisers

For boosters, Kenya produces not only the lowest values throughout but also displays a fairly even distribution which shows that the low averages are not due to individual texts with extreme values. In general, every category contains several outliers that have significantly more boosters than the majority of texts, the most outstanding one occurring in the U.S. lifestyle domain. A look at the headline of the text is enough to explain this, as it treats a topic based on comparisons and therefore includes 13 hits of *more* in a total of just 303 word tokens (4.3%):

> [13] Men obsessed with their muscles are **more** sexist, **more** likely to believe thin women are **more** attractive (Goldwert 2012, *Daily News*; emphasis added)

With regard to minimisers, represented in the lower plot, it can be seen that five categories display a median value of 0, three of which in the domain of hard news. In all domains except sports, the USA and the UK have the highest median values, while especially in economy and politics Australia and Hong Kong are very similar. Lifestyle articles show the highest, but also most constant medians whereas sports news contains the highest internal variation. The most dominant outlier concerning minimisers originates from the political news section from the UK however, and is a 304-word-article dealing with election results, demonstrating yet again that some topics, like the one mentioned above, can bring with them larger clusters of these lexical items:

> [14] […] Collins took 4,648 votes, **nearly** 22 per cent of those cast […] With **just** 451 votes, Mr Beckett was **just** 190 votes ahead […] Mr McDonald, who won a share of **just** over 60%, said the result sent the message to the Coalition […] Ukip's Richard Elvin came second with **just** under 12%. (Rojas 2012, *The Daily Telegraph*; emphasis added)

Looking at the meaning of these results in terms of evaluative language and social distance, a clear trend can be observed for lifestyle articles to feature significantly many boosters and minimisers and therefore make use of a more evaluative style. The opposite is true for hard news, which is not unexpected as evaluative language again clashes with the fact-focused structure of these texts. News stories like accidents, catastrophes or crimes do not need to be boosted or minimised but often, especially when deaths are included, speak for themselves. In this aspect, there are no significant differences between the hard news of more quality and more popularly oriented papers – in the former, boosters and minimisers make up 0.55%, in the latter it is 0.53%, which results in a χ^2-value of 1.03 (df=1, p=0.31). The same holds true applied to the whole corpus, where popular papers display 0.77% of boosters and minimisers in contrast to 0.76% in quality papers, with χ^2=1.86 (df=1, p=0.17).[25]

With regard to the varieties, the strongest differences exist between the news from Kenya and the UK. In general, the UK and Australia contain most boosters and minimisers, and therefore seem to evaluate most strongly. The significance tests among the regions showed that especially the low numbers in Kenya stand out, which means that here, evaluative language in terms of boosters or minimisers is rare. Hong Kong features the second lowest number of these items, which indicates once more that the news is more standardised and aim for objectivity more than in the native varieties. This is reflected also in the internal ranges, for which Hong Kong produces a difference between its domains of just 0.33 percentage points, in contrast to the USA (0.41), Kenya (0.43), the UK (0.46) and Australia (0.57) (see appendix 4 for a detailed overview of the ranges for all features).

Evaluative language – Adjectives and adverbs

The third feature to contribute to the parameter of evaluative language is the use of strongly connoted adjectives and adverbs. Connotations are of course highly subjective, which makes this feature very prone to misinterpretation. As was explained in section 3.2.2.3, the queries here included only the most obvious items; this means that many potential candidates for this linguistic feature are left out, yet it minimises the subjective influence as best as possible and was therefore preferred.

In order to determine the final list of lemmata for the corpus queries, a list of the 100 most frequent adjectives and lexical adverbs (part of speech-tags JJ, JJR, JJS, RB, RBR and RBS) was generated for every variety. Many items in these lists of course overlapped; in total they produced 145 different lemmata, which is 9.86% of all lemmata with these part of speech-tags in the corpus. In terms of frequencies, their cumulative occurrences are responsible for 53.36% of all adjectives and lexical adverbs. These 145 items were filtered manually for those with the clearest connotations, which led to the following lemmata for the final search:

[25] As *more* and *most* are the dominant terms among the ones queried here and are common throughout all papers and domains due to their function to construct comparatives and superlatives, the significance was tested again excluding these two terms. The difference between quality and popular papers remains insignificant however, at χ^2=3.08 (df=1, p=0.0792).

Table 5.12: Frequencies of positive and negative adjectives and adverbs

Lemma	Freq.	%	Lemma	Freq.	%
bad	658	0.03	*important*	553	0.03
dead	296	0.01	*low*	804	0.04
difficult	379	0.02	*poor*	292	0.01
easy	397	0.02	*right*	393	0.02
free	404	0.02	*small*	784	0.04
good	2,936	0.14	*special*	330	0.02
great	998	0.05	*strong*	657	0.03
happy	343	0.02	*top*	755	0.04
hard	633	0.03	*well*	1,879	0.09
		∑	**13,491**	**0.63%**	

Although some of these items appear clearly negative or positive, a categorisation was avoided as several of the lemmata do not obviously tend towards one or the other, but depend on the context. Furthermore, instances of these words preceded by *no*, *not* or *n't* were not excluded in this search; in contrast to minimisers or boosters, where the respective function can be neutralised by a negation as was seen in the given examples, the evaluative content of the above listed lemmata does not change. It can be weakened or turned to the opposite, yet it remains an instance of evaluative language:

[15] Mr O'Connell said he knew the stretch of road where the fatal accident happened "reasonably well". "My understanding of it is that stretch of road is **not bad**," he said. (Jabour 2012, *The Sydney Morning Herald*; emphasis added)

[16] Without Murray the prospects for the British team are **not good**, though they do have the home advantage with the tie being held in the Ricoh Arena in Coventry. (Lee 2013, *Daily Mail*; emphasis added)

In both examples, very clearly connoted adjectives are negated, nevertheless they are preferred to simply using their opposites. The politician in the first excerpt does not say that the road was good, but just that it was "not bad", a preference which makes the statement vaguer and still emphasises the negative rather than the positive aspects. The second excerpt demonstrates this the other way around; the chances for the team in question are said to be "not good", but the addition afterwards, the home advantage, indicates that they are not bad either and there is reason to be hopeful, which is strengthened by including the word *good* rather than *bad*, even if it is negated. Negated instances were therefore included in the analysis as they do not lose their evaluative function and contribute to the degree of evaluation of the whole text. In total, the queries produced 164 instances of negated forms, most prominently in connection with *good* (40), *easy* (37) and *happy* (14).

As was the case in previous analyses, the distribution of these markers of evaluation reveal bigger differences between the domains than between the varieties. Table 5.13 shows the absolute and relative frequencies within every category as well as for the varieties and domains.

Table 5.13: Frequencies of evaluative adjectives and adverbs

	AUS	HK	KEY	UK	USA	∑ *% of domain*	*range*
Economy	557 0.78%	470 0.57%	335 0.51%	615 0.73%	519 0.61%	2,496 0.64%	280 0.27
Hard News	159 0.29%	254 0.39%	171 0.29%	397 0.48%	292 0.34%	1,273 0.37%	238 0.19
Lifestyle	845 0.83%	804 0.68%	771 0.61%	866 0.76%	765 0.65%	4,051 0.7%	101 0.22
Politics	370 0.48%	426 0.52%	268 0.36%	457 0.52%	409 0.46%	1,930 0.47%	189 0.16
Sports	785 0.98%	673 0.78%	470 0.73%	923 0.92%	890 0.91%	3,741 0.87%	453 0.25
∑ *% of variety*	2,716 0.71%	2,627 0.61%	2,015 0.52%	3,258 0.69%	2,875 0.61%	13,491 0.63%	
range	686 0.69	550 0.39	600 0.44	526 0.44	598 0.57		764 0.69

Among the varieties, Australia and the UK produce the highest values in total, while Kenya shows by far the lowest. This is in line with what was found for minimisers and boosters, where the ranking of the varieties in terms of frequencies was the same. Here, the differences are significant with a cumulative χ^2-value of 156.34 (df=4, p<0.000), to which the low numbers in Kenya contribute 75.98 and the high numbers in Australia and the UK 39.6 and 33.36 respectively. Furthermore, every variety displays the same internal ranking of the domains: in all regions, hard news produces the lowest values, politics the second lowest, economy forms the middle, lifestyle shows the second highest results and sports contains most hits.

The overall differences between the domains are much more significant still than those between the varieties at χ^2=1005.28 (df=4, p<0.000), and the values for the low numbers in hard news (χ^2=370.07) and politics (χ^2=163.67) as well as the high frequencies in sports news (χ^2=411.95) confirm their outstanding roles. The distributions within the individual categories follow the same trend; although all categories produce a number of outliers, the domains of sports, lifestyle and also economy show very wide ranges, thereby proving that their higher numbers are not due to few texts, whereas hard news and politics are more compact.

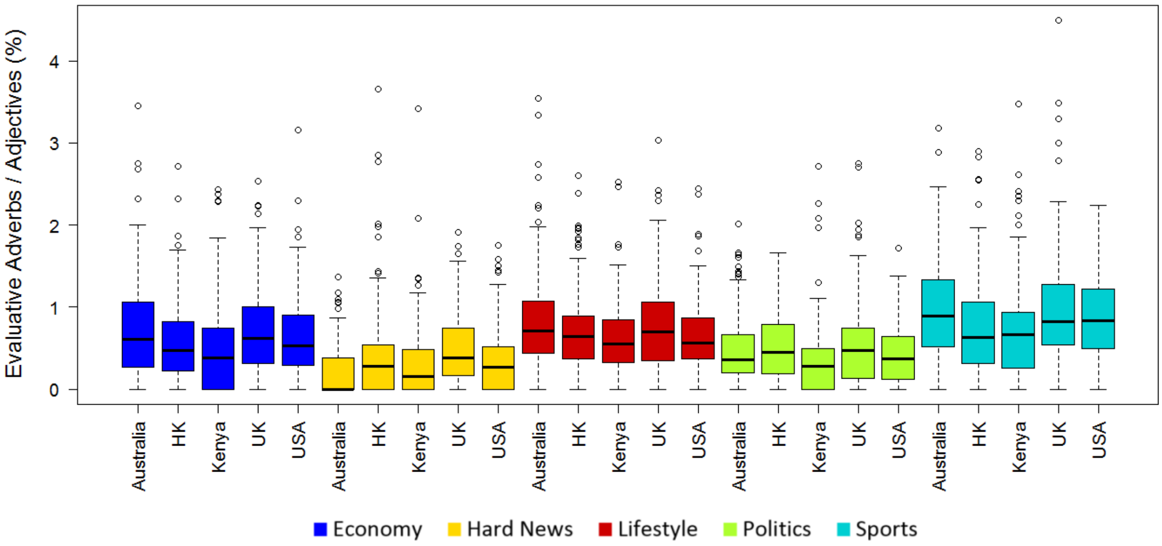


Figure 5.21: Distribution of evaluative adjectives and adverbs

Regarding the domains, the results for this linguistic feature are thus also comparable to the insights gained from the analysis of minimisers and boosters. Political and hard news appear to express evaluation much less than the other domains, and especially sports news and lifestyle articles make frequent use of this means, which again can be assumed to originate in their respective topics. With regard to different newspaper types, a significant difference can be found between more popular and quality papers. The latter contain fewer instances of the evaluative adjectives and adverbs queried here, with χ^2=26.77 (df=1, p<0.000), a difference which is particularly expressed in the domains of sports (χ^2=14.3, df=1, p=0.0002) and hard news (χ^2=12.45, df=1, p=0.0004).

Evaluative language – Modality

The last linguistics indicator to be analysed here as reflecting evaluative language is modality. In contrast to the previous ones, it is summarised under one part of speech-tag and can therefore be more easily queried and quantified. The results were already presented in Figures 5.14 and 5.15 under social role relationship (section 5.2.1), yet the focus connected to this feature here is slightly different. Fowler (1991, 85-87) defines four meanings that modality can have in evaluative terms, namely truth, obligation, permission and desirability. In many instances, these categories overlap and since Fowler uses qualitatively analysed examples and includes other items of modality than verbs, it is problematic to apply this model to a corpus-based study. Many of the modal verbs analysed here can be assigned to different categories depending on their context, as the following examples demonstrate:

[17] Facebook's ill-fated shareholders **can** take solace in the fact that they probably didn't make as big a wager on the social network's stock as the owners of Instagram. (Jaccarino 2012, *Daily Mail*; emphasis added)

[18] "I'm out of contract this summer and have been told I **can** leave [...]" (Lewis 2013, *The Sun*; emphasis added)

[19] Kenol is targeting to sell a significant stake in two to three years when the company is expected to be profitable and it **will** invite multiple bids. (Juma 2013, *Daily Nation*; emphasis added)

[20] "As a result of Ethan's injuries he became eligible to donate his organs, and this **will** hopefully prevent another family or families from suffering [...]". (Flower 2011, *Herald Sun*; emphasis added)

In the first excerpt, *can* is used in the sense of *being able to* and would thus fit best into the category of truth, whereas in example [18], it shifts towards a permission. *Will* in the third sentence is used as a marker for truth as it underlines what is going to happen with more certainty than for instance *might* would in this case. In example [20] however, *will* is conditioned by the adverb *hopefully*, making it a desire rather than a truth. Other problematic cases are *may* and *might*, which can signal permission as well as a category that would have to be called probability. Across these items, there are no indicators as to their exact use that could be queried on the basis of part of speech-tagging or lemmatisation; a reliable investigation and interpretation of these and other potential categories would require an in-depth analysis of the texts in question, and will therefore not be included here.

Nevertheless, even without concrete meanings implied by individual modals, modality reflects evaluative language in the sense that via this feature, the author of a news item can express a subjective attitude. It was seen in section 5.2.1 that among the varieties, Hong Kong and the USA have the lowest values for modals with 1.09% and 1.07% respectively, in contrast to Australia and the UK, which produce higher frequencies (1.31% and 1.3%). The variation between the regions is significant at χ^2=196.39 (df=4, p<0.000). Regarding the domains, the differences are more expressed with a cumulative χ^2-value of 538.83 (df=4, p<0.000), and especially hard news and politics stand out as containing fewer and more modals respectively than statistically expected.

In the context of the subdimension of social role relationship, these results were related to factual knowledge about the topic and the effect of modals to express insecurity or vagueness as well as speculation. These features of modality of course have evaluative properties as well, and cannot be assigned to either social role relationship or social distance, but can reflect either one or both depending on how they are employed. For Hong Kong, the low numbers of modals are in line with what has been found for the previous indicators of evaluative language, namely that news from that region tends towards a lower degree of evaluation. For the USA on the other hand, the previous two features already revealed a mixed picture, and a clear trend cannot be determined yet. For Australia and the UK, which contain most modals, the cases are similar; while Australia displays a rather straight tendency towards evaluative style, the UK showed rather few boosters and minimisers and thus appears more diverse.

The same can be observed among the domains as well, where especially the high frequencies of modals in political news contradict the previous results, where this domain appeared less evaluative than the others. The results for hard news on the other hand are much more consistent, as they display comparatively low evaluative content in all markers analysed here.

1st and 2nd person pronouns

Apart from forms of evaluative language, the subdimension of social distance is reflected by other linguistic features which construct closeness or distance between the discourse participants. One of these features is the use of pronouns, but in contrast to the analysis of this feature for the parameter of goal orientation (section 5.1.2), which focused on 3rd person pronouns, 1st and 2nd person pronouns are of interest at this point as they reflect the relationship between reader and author. Again, all pronouns were queried including their possessive forms, taking into account the words *I*, *mine*, *me*, *myself*, *my*, *you*, *your*, *yours*, *yourself*, *yourselves*, *we*, *us*, *our*, *ours* and *ourselves*. Table 5.14 shows the cumulative absolute and relative frequencies of these items per category.

The distributions illustrated in this table confirm what was found for instances of evaluative language with regard to the varieties, namely that again, the news items from Kenya and Hong Kong show general trends towards more distance. Concerning pronouns however, the USA also contains fewer instances, while the UK and Australia feature more. The differences are highly significant with a combined χ^2-value of

1914.13 (df=4, p<0.000), with the low numbers in Hong Kong at χ^2=1041.66 and the high numbers in the UK and Australia at χ^2=467.35 and χ^2=341.78 respectively.

Looking at domains, the tendencies suggested in the previous analyses are strengthened regarding lifestyle and sports news. They display by far the highest numbers of 1st and 2nd person pronouns, thus aiming most directly at their readers. Among the other three domains however, the roles are slightly different than before, with economy displaying the lowest numbers in all regions, followed by hard news. Again, the differences are reflected in high levels of significance (χ^2=8070.24, df=4, p<0.000), which is dominantly caused by the domains of economy (χ^2=2026.55) and lifestyle (χ^2=3595.54).

Table 5.14: Frequencies of 1st and 2nd person pronouns[26]

	AUS	HK	KEY	UK	USA	∑ *% of domain*	*range*
Economy	586	278	468	678	541	2,551	400
	0.82%	0.34%	0.71%	0.8%	0.63%	0.66%	0.48
Hard News	537	247	457	1,093	817	3,151	846
	0.98%	0.38%	0.78%	1.32%	0.96%	0.91%	0.94
Lifestyle	3,154	1,921	3,314	3,629	2,660	14,678	1,708
	3.11%	1.63%	2.62%	3.18%	2.26%	2.54%	1.55
Politics	1,199	406	649	1,200	851	4,305	794
	1.57%	0.49%	0.87%	1.36%	0.96%	1.05%	1.08
Sports	1,937	1,246	1,028	2,568	2,067	8,846	1,540
	2.42%	1.44%	1.59%	2.57%	2.12%	2.06%	1.13
∑	7,413	4,098	5,916	9,168	6,936	33,531	
% of variety	1.93%	0.95%	1.52%	1.95%	1.46%	1.56%	
range	2,617	1,674	2,857	2,951	2,119		3,382
	2.29	1.29	1.91	2.38	1.63		2.84

The domains and varieties do not just display variation in terms of cumulative frequencies of the pronouns however, but also with regard to the distribution of the different pronouns. Figure 5.22 displays the three categories of 1st person singular, 1st person plural and 2nd person, for which singular and plural were combined under the lemma *you* and its various forms, in percentages of the respective categories.

[26] The discrepancies between Table 5.5 in section 5.1.2 and Table 5.14 are due to the personal pronoun *one* (lemmata *one* and *oneself*), which was counted for neither subdimension. It will be part of the analysis of pronominal use for medium (section 5.3.1).

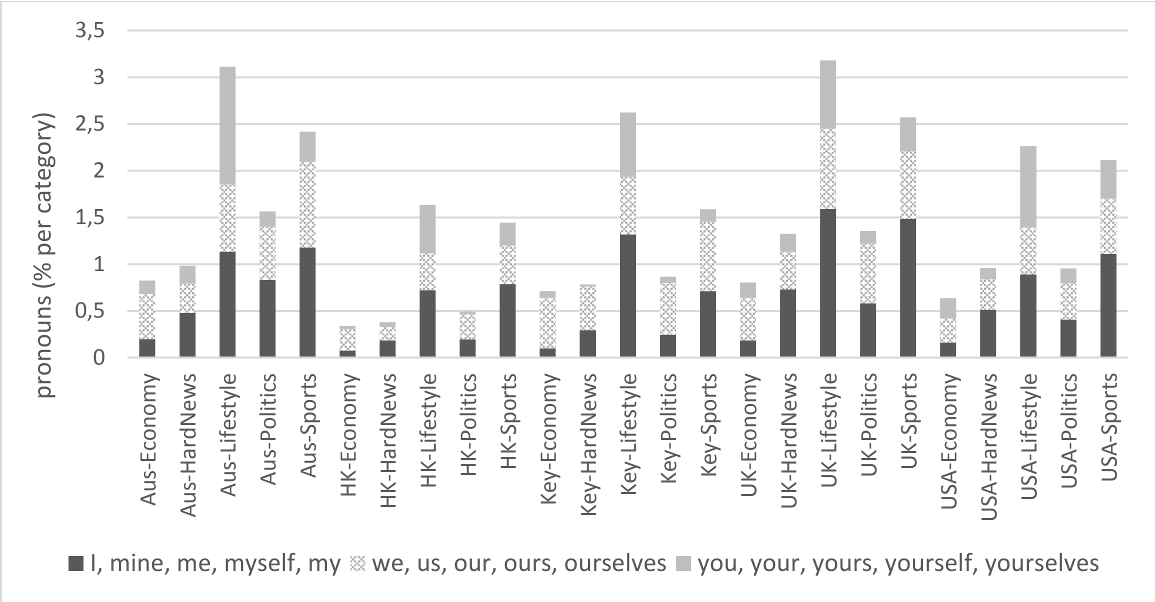


Figure 5.22: Relative frequencies of different pronouns

Of these three categories, all five varieties contain most instances of the 1st plural singular forms and least occurrences of the 2nd person pronouns. Within the different domains however, slight discrepancies can be found from this overall trend.

The 1st person singular pronoun is dominant in sports news across all regions, and strong also in lifestyle news, where only in Australia the 2nd person pronoun is more frequent, and in hard news, where the only exception is Kenya. In economic news on the other hand, the 1st person plural pronoun is most frequent in all cases. Political news displays a more even relation between these two, and only in Australia a clear tendency towards the 1st person singular can be observed.

These results suggest that the domains include also varying numbers of direct quotations, which have to be assumed as a frequent source of occurrences of *I*, *mine*, *me*, *my* and *myself*. The dominance of the variations of *we* in economic news indicate that here news actors, when quoted directly, refer rather to a company than to themselves, thus shifting from a personal to a professional identity (see Mieroop 2005 for a discussion of different speaker identities). The 2nd person pronouns are most strongly represented in lifestyle articles, which is compatible with previous results, especially regarding mood, indicating that these news items are more geared towards the reader and the interpersonal aspect is emphasised more strongly.

Among the varieties, as was already said above, Australia and the UK appear to address their readers directly most often. This again confirms the results concerning interrogatives and imperatives, where these two varieties also produced the highest frequencies. In the light of these as well as the previous results for this subdimension, it can thus be concluded that the news articles from these regions aim at a lower distance between authors and readers. The opposite is true for Hong Kong and Kenya, although in the case of the latter, the results for mood were less outstanding. The USA however produces very mixed results; comparatively high frequencies of interrogatives as well as minimisers and boosters indicated a strong reader orientation, yet regarding evaluative language in terms of strongly connoted adverbs and adjectives and the use of 1st and 2nd person pronouns, it appears more similar to the varieties from Kenya and Hong Kong.

Place references

Another way for journalists to construct closeness to their readers is to refer to a shared environment. Via these references, the relevance of the place for both parties is emphasised and the news worthiness raised. In turn it can be assumed that places that feature frequently are perceived as culturally or geographically near.

As was said in section 3.2.2.3, the filtering of names is problematic in the dataset as the part of speech-tagger does not differentiate between personal names or names of places, companies, products or other entities. In contrast to personal names however, the group of place names is a fairly closed field and the dominant items make up the vast majority of the whole group. Furthermore, place names are less ambiguous and more reliably tagged. For these reasons, the 20 most frequent place names per variety were filtered out manually and are shown, together with their absolute and relative frequencies in their respective region, in Table 5.15. The numbers for the items *UK, USA, NSW* and *WA* include their respective long and alternative forms. *EU* on the other hand was not combined with the counts for *Europe*, as their referents are not identical.

As was to be expected after looking at the keyword lists in section 5.1.1, every variety features place names from its own region very frequently. This trend is most clearly expressed in Australia and the USA, where eleven and ten respectively of the top twenty names are within the country. This reflects what was discussed in section 2.3.3, that because of the size of these countries news focuses more on local and regional events. The UK includes seven national place names among its top twenty, which in this case can be related to its individual parts, of which Scotland and Wales occur in the list, as well as to the overlapping concepts of the UK and Britain. Hong

Kong and Kenya on the other hand feature fewer national names with five each,[27] counting Chinese place names as national for Hong Kong.

Table 5.15: Top 20 place references per variety

No	Word	Freq.	%	No	Word	Freq.	%
Australia							
No	Word	Freq.	%	No	Word	Freq.	%
1	Australia	716	0.19	11	NSW	77	0.02
2	US	289	0.08	12	England	74	0.02
3	Sydney	278	0.07	13	Adelaide	63	0.02
4	WA	262	0.07	14	New Zealand	60	0.02
5	Melbourne	242	0.06	15	Japan	58	0.02
6	Queensland	233	0.06	16	India	51	0.01
7	Perth	192	0.05	17	World	47	0.01
8	Brisbane	178	0.05	18	Asia	43	0.01
9	China	113	0.03	19	Canberra	43	0.01
10	Victoria	97	0.03	20	Europe	41	0.01
			∑	**3,157**	**0.82%**		
Hong Kong							
No	Word	Freq.	%	No	Word	Freq.	%
1	China	1,598	0.37	11	Korea	128	0.03
2	Hong Kong	789	0.18	12	Singapore	120	0.03
3	US	535	0.12	13	Taiwan	109	0.03
4	Beijing	377	0.09	14	London	93	0.02
5	Japan	276	0.06	15	Germany	87	0.02
6	Shanghai	175	0.04	16	EU	86	0.02
7	Asia	170	0.04	17	Tokyo	78	0.02
8	India	166	0.04	18	Guangdong	75	0.02
9	World	149	0.03	19	France	68	0.02
10	Australia	141	0.03	20	Thailand	61	0.01
			∑	**5,281**	**1.22%**		
Kenya							
No	Word	Freq.	%	No	Word	Freq.	%
1	Kenya	1,432	0.37	11	Rwanda	47	0.01
2	Nairobi	466	0.12	12	Nigeria	40	0.01
3	Africa	293	0.08	13	London	39	0.01
4	Mombasa	194	0.05	14	China	37	0.01
5	World	151	0.04	15	Sudan	37	0.01
6	Uganda	101	0.03	16	India	36	0.01
7	US	88	0.02	17	France	34	0.01
8	Tanzania	84	0.02	18	UK	32	0.01
9	Westgate	64	0.02	19	Turkana	29	0.01
10	Rome	48	0.01	20	Europe	26	0.01
			∑	**3,278**	**0.84%**		
UK							
No	Word	Freq.	%	No	Word	Freq.	%
1	Britain	434	0.09	11	Germany	93	0.02
2	UK	380	0.08	12	Manchester	88	0.02
3	London	354	0.08	13	America	87	0.02
4	England	351	0.07	14	Scotland	84	0.02
5	US	339	0.07	15	China	72	0.02

[27] *Westgate* in the Kenyan list refers to the Westgate Shopping Centre in Nairobi, which is mentioned frequently as it was the scene of a mass shooting in 2013.

6	EU	206	0.04	16	Japan	68	0.01
7	Europe	162	0.03	17	Italy	65	0.01
8	World	122	0.03	18	Spain	65	0.01
9	France	120	0.03	19	Wales	64	0.01
10	Russia	105	0.02	20	Africa	60	0.01
		∑ 3,319	**0.71%**				
USA							
No	Word	Freq.	%	No	Word	Freq.	%
1	US	670	0.14	11	Texas	89	0.02
2	New York	345	0.07	12	Manhattan	87	0.02
3	Boston	224	0.05	13	America	75	0.02
4	Europe	208	0.04	14	Brooklyn	73	0.02
5	Italy	196	0.04	15	Japan	68	0.01
6	Germany	148	0.03	16	London	68	0.01
7	Washington	132	0.03	17	World	68	0.01
8	France	128	0.03	18	Georgia	67	0.01
9	China	121	0.03	19	Greece	64	0.01
10	EU	115	0.02	20	California	63	0.01
		∑ 3,009	**0.63%**				

In general, this trend towards a regional focus is of course to be explained by the geographical nearness of the places, strengthening the news value of proximity, as well as the relevance that national events hold for the population of a country, which contributes to the news value of impact (Bednarek and Caple 2014, 155-56). Concerning the scope of the regions, it can be seen that Hong Kong, Kenya and the UK make frequent references to countries on their continent, whereas the USA and Australia contain very few of these. In the case of Australia, this is due to the focus being so dominantly on regional events, but also due to its relative isolation. New Zealand is referred to 60 times and apart from this, the closest references are the U.S. and various Asian countries. Especially China and Japan occur in the other regions' lists as well, however, both in the UK and the USA more than half of their cumulative hits, 61.43% and 57.67% respectively, originate from the domain of economy, showing that the interest is concentrated on this area. In Australia on the other hand the hits spread more evenly over economy (70 hits, 40.94%) and politics (67 hits, 39.18%), and even lifestyle contributes 21 occurrences (12.28%).

In the case of the USA, the rare references to other American places contradicts what would be expected on the basis of closeness. *America* features 75 times in the US-part, yet only four of these refer to Latin America, and another two to South America. The first reference to a non-US place on the American continent is *Toronto* with 46 hits, followed by *Mexico* (38 hits) and *Canada* (27 hits). Instead, among the twenty most frequent items we can find seven references to places in Europe. A curious case here is *Italy*, which reaches by far more hits in the USA than in any other variety; a check for this item shows that it occurs most frequently in hard news and politics, where it clusters in rather few texts and in almost all cases is connected to law suits against Berlusconi, whose name also occurred in the US keyword lists. It can therefore not be said that there is a general relevance of this country for US news, but that its frequent mentioning is due to the political scandals happening at the time.

Nevertheless, Europe clearly appears as a much stronger reference point for the USA than its immediate neighbouring countries. Again, the dominant domains for these references are politics and economy, which contain 83.5% of the cumulative European place hits (88.1% when excluding *Italy*). This reflects a more general tendency of the distribution of national and international pace references; in hard news, national references are significantly more common than international ones in every variety, and the same holds true for lifestyle articles, with the exception of Kenya. In politics and economy on the other hand international references are more relevant, whereas sports news displays different tendencies across the corpus (see also Fest, forthcoming). Table 5.16 lists the absolute frequencies of national and international hits per category. As the numbers of national and international items differ within and also across the regions, percentages are not included; instead the significance was tested within every category and the values which are lower than statistically expected are given in brackets.

Table 5.16: National versus international place references

		AUS	HK	KEY	UK	USA
ECO	national	(404)	947	(551)	460	(360)
	international	289	(503)	377	(402)	371
	χ^2-value	140.42	55.37	30.88	0.11	52.13
	p-value	<0.000	<0.000	<0.000	0.739	<0.000
HN	national	268	326	293	267	450
	international	(27)	(208)	(22)	(79)	(58)
	χ^2-value	41.78	3.83	108.66	91.47	200.51
	p-value	<0.000	0.05	<0.000	<0.000	<0.000
LIFE	national	381	560	(412)	193	326
	international	(97)	(327)	238	(111)	(97)
	χ^2-value	5.59	15.99	3.91	15.12	55.93
	p-value	0.018	0.0001	0.048	0.0001	<0.000
POL	national	660	(725)	353	(356)	(349)
	international	(127)	617	(127)	603	588
	χ^2-value	40.31	6.83	11.99	134.36	310.98
	p-value	<0.000	0.009	0.0005	<0.000	<0.000

SPO	national	(668)	(456)	(576)	479	336
	international	236	612	329	(369)	(70)
	χ^2-value	1.59	112.93	5.1	5.95	96.54
	p-value	0.207	<0.000	0.024	0.0147	<0.000
Σ	**national**	**2,381**	**3,014**	**2,185**	**1,755**	**1,821**
	international	**776**	**2,267**	**1,093**	**1,564**	**1,184**

In politics, Australia's focus on internal matters is once more expressed in the significance of national references, which stands in contrast especially to the UK, the USA and Hong Kong, while Kenya displays a less significant difference in this domain. In economic news, Hong Kong is the outstanding variety as here, national events appear much more relevant than international ones. However, as Hong Kong is an important business centre and English has long been regarded as a gateway to professional success (see section 2.1.3), it is logical that English-language press concentrates on these aspects.

Regarding the degree of social distance constructed in the different domains and varieties, it can be said that especially Australian, but also US-American news items have a clear focus on internal matters and work on the level of states rather than the whole countries. With these references, target groups from the respective areas are addressed specifically while in the other varieties, the national level is more dominant. Regarding external references, geographical closeness is a decisive factor for all varieties except the USA, which appears closer to Europe than to other American countries. The distance towards these regions is thereby constructed as fairly low.

With regard to the domains, the construction of national in-groups is particularly emphasised in hard news and lifestyle. In these domains, events within the countries are clearly more important than elsewhere, and international relations are of little consequence. In economic and political news in turn these dimensions are highly relevant, and readers are brought into contact with other countries and societies more frequently. This extended scope increases the closeness towards the respective regions abroad, but also the distance towards the readers as the relations described here are less intimate and put the assumed shared background on a much broader scale, for instance of the same continent or political alliance rather than of the same city or county. Economic and political news articles require this wider focus for their topics as international relations play a major role here, and therefore automatically imply a higher distance than the other domains.

Contractions

In contrast to the analysis of place references, which again relied on manual sorting, contractions are very straightforward and can be queried on the basis of the corpus annotation. In total, seven contracted forms were counted: *n't, 's* (excluding possessive forms)*, 're, 'm, 've, 'll* and *'d* (cf. Biber et al. 1999, 1128-32). All of these forms are recognised by the tagger as separate words, an item such as *don't* is thus counted as the two words *do* and *n't* (*not*). Table 5.17 lists the absolute frequencies of all contractions as well as the percentage they make up of all word tokens within every category.

Table 5.17: Frequencies of contractions

	AUS	HK	KEY	UK	USA	∑ *% of domain*	*range*
Economy	365	201	50	283	475	1,374	425
	0.51%	0.25%	0.08%	0.33%	0.56%	0.35%	0.48
Hard News	265	183	37	380	507	1,372	470
	0.48%	0.28%	0.06%	0.46%	0.52%	0.4%	0.46
Lifestyle	1,466	862	363	1,366	1,618	5,675	1,255
	1.45%	0.73%	0.29%	1.2%	1.9%	0.98%	1.61
Politics	398	180	55	440	691	1,764	636
	0.52%	0.22%	0.07%	0.5%	0.59%	0.43%	0.52
Sports	1,341	735	145	1,157	1,674	5,052	1,529
	1.67%	0.85%	0.22%	1.16%	1.88%	1.18%	1.66
∑	3,835	2,161	650	3,626	4,965	15,237	
% of variety	1%	0.5%	0.17%	0.77%	1.05%	0.71%	
range	1,201	682	326	1,083	1,199		1,637
	1.19	0.63	0.23	0.87	1.38		1.84

The results displayed here very much strengthen the tendencies shown for previous features that news from Kenya and Hong Kong constructs a higher distance between author and reader than the other varieties. The differences between the regions with regard to contractions are significant at χ^2=3154.44 (df=4, p<0.000), and both L2 varieties display lower values than statistically expected, whereas the other three are all above their expected values. Furthermore, the varieties produce the exact same ranking in every domain: the USA always shows the highest values, followed, in that order, by Australia, the UK, Hong Kong and Kenya.

Regarding the domains, the differences that can be observed again set news from sports and lifestyle apart from the rest. The two domains show by far the highest frequencies of contractions, whereas economy, politics and hard news items make use of them only rarely. The variation in this dimension is significant at χ^2=3597.59 (df=4, p<0.000), with lifestyle and sports having higher, the other domains lower values than expected. These tendencies are reflected also on the level of individual texts (see Figure 5.23), where it is confirmed that the averages presented in Table 5.17 are not the results of individual extremes, but of fairly even distributions.

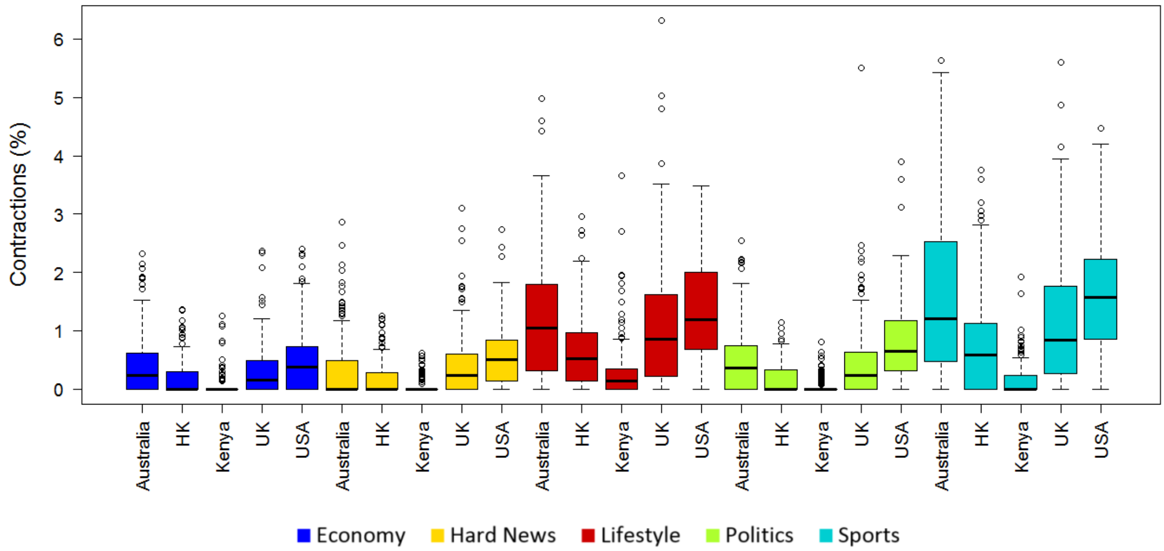


Figure 5.23: Distribution of contractions

Lifestyle and sports again stand out immediately, and these domains also produce fewest outliers. The broadest range is observable in sports news from Australia, which contains only one outlier in total. The other domains produce more exceptional values, which is due to the very restricted range of the bulk of the texts; as can be seen, many of the plotted outliers are not far from the upper whisker and are therefore not much apart from the core values. In the three domains of economy, politics and hard news, Australia, Hong Kong and Kenya contribute only few and moderate outliers, whereas the more extreme values occur in the parts from the USA and the UK.

The frequent use of contractions in language is an indicator for a more familiar, colloquial style, higher numbers thus reflect a lower distance between the discourse participants (Steiner 2004, 18, 35; Neumann 2013, 67-71). For the varieties and domains analysed here, the results and trends are very clear and, when seen in the context of the analyses of the previous sections, little surprising. Lifestyle and sports articles have been shown before to focus on an interpersonal more than a factual dimension, which is compatible with a lower distance. The domains of economy, hard news and politics, with their more formal and serious topics, tend towards the opposite and thus do not feature more contractions.

Regarding the varieties, the results for Kenya and Hong Kong also confirm previous numbers. The news from both varieties construct a high distance between authors and readers and also show little internal variation in this respect, as the ranges in Table 5.17 illustrate. This again points towards a stronger standardisation of news and a more formal or heavily controlled approach to journalism. Furthermore, it hints at an orientation of these varieties along written norms, as contractions are a characteristic feature of spoken language. Although this aspect is connected to social distance, it will not be discussed further here; instead, it will be taken up again for the subdimension of medium, where it will be put into context with other markers of spoken discourse.

Voice

The linguistic feature of voice is a very specific one when analysing news. In general, a high frequency of passive voice indicates a focus on events rather than actors, creating a higher distance (Lamb 1991, 255-56; Neumann 2013, 127), yet in news language, it has to be considered that journalistic guidelines emphasise the use of active voice as a means to make the story more vital and easier to process for an audience who does not have much time (Harcup 2005, 107, 129; Pape and Featherstone 2005, 42, 51; Keeble 2006, 96). A strong dominance of active voice can therefore be expected, and is reflected in the results. For the present investigation, passive voice was queried as every instance of the following command:

> *[lemma="be"][pos!="SENT|:"]?[pos!="SENT|:"]?[pos="VBN"]*

The first element defines a form of the verb *be*, the last a past participle. Since the second and third element are optional, this alone filtered out all simple passive constructions as given in example [21]. The middle elements were added to account for passive constructions with additional modifiers, as demonstrated in examples [22], [23] and [24].

[21] This week billions of pounds **was wiped** off London share values on Black Thursday. (Hendry 2007, *The Sun*; emphasis added)

[22] The decision **was hotly contested** by West Ham players who roughed up the referee. (Sheriff 2013, *Daily Nation*; emphasis added)

[23] The store owner **was not injured** during the incident. (Orr 2012, *The Sydney Morning Herald*; emphasis added)

[24] The identity of the other two victims **is not yet known** [...] (Robertson 2013, *The Sydney Morning Herald*; emphasis added)

Table 5.18 summarises the raw frequencies of passive constructions and their relative frequencies in relation to all constructions including a finite verb per category.

Table 5.18: Frequencies of passive voice

	AUS	HK	KEY	UK	USA	∑ *% of domain*	*range*
Economy	644 10.07%	585 8.75%	760 13.96%	828 10.5%	563 7.33%	3,380 9.91%	265 6.63
Hard News	1,052 18.39%	1,053 16.92%	1,223 20.19%	1,674 19.01%	1,291 14.1%	6,293 17.5%	622 6.09
Lifestyle	869 8.2%	951 8.59%	1,509 11.42%	1,169 9.58%	887 7.52%	5,385 9.15%	640 3.9
Politics	921 12.37%	861 12.02%	1,071 14.83%	1,004 11.77%	728 8.52%	4,585 11.78%	343 6.31
Sports	730 8.72%	703 8.61%	578 9.91%	859 8.12%	661 6.38%	3,531 8.15%	281 3.53
∑ *% of variety*	4,216 10.94%	4,153 10.56%	5,141 13.61%	5,543 11.53%	4,130 8.69%	23,174 10.98%	
range	408 10.19	468 8.33	931 10.28	846 10.89	728 7.72		1,111 13.81

The patterns observable here are not quite as homogenous as the ones for the use of contractions, yet the extreme values are throughout all domains produced by Kenya for the most and by the USA for the fewest passives. The differences between the varieties are significant at χ^2=545 (df=4, p<0.000), to which Kenya and the USA alone add 238.94 and 226.7 respectively. Looking at the domains, the differences become even more expressed. Hard news contains by far the highest amount of passive constructions in every variety, followed, also in all regions, by politics. The other domains are more similar to each other and display a slightly different status depending on their variety. Based on the overall averages, the domain-related variation is significant at χ^2=2187.03 (df=4, p<0.000). An analysis across all texts shows that the majority of values is evenly distributed.

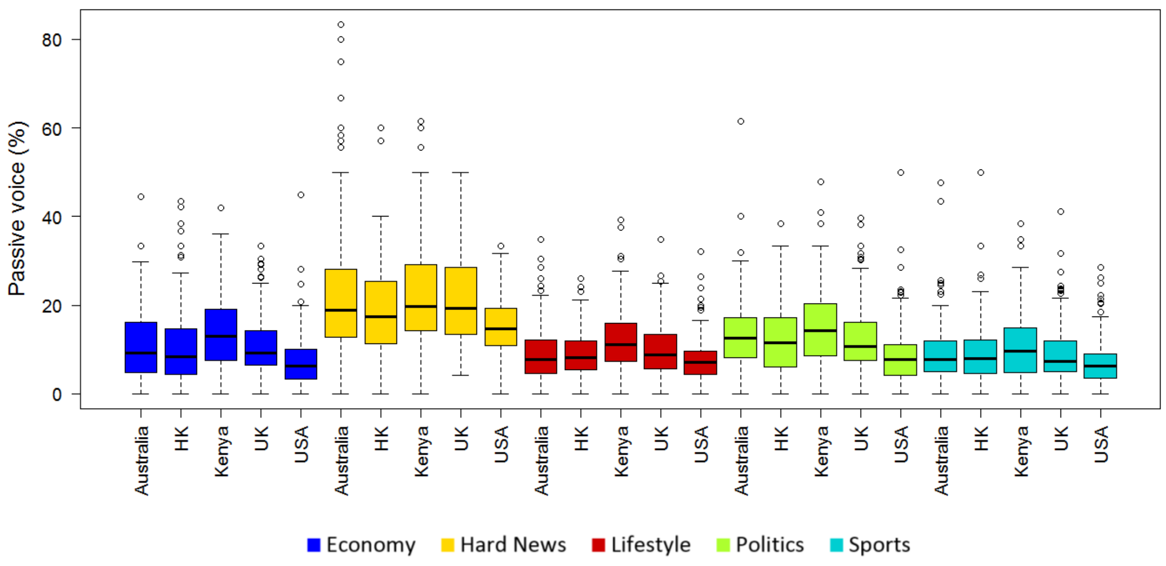


Figure 5.24: Distribution of passive voice

As was suggested already by the average values in Table 5.18, the USA and Kenya stand out whereas the other three varieties hardly differ at all with regard to their median values and only little regarding the range covered by their respective boxes and whiskers. Hard news and politics, the domains with the highest average amounts of passive voice, also display the widest ranges, proving that the averages were not skewed by individual texts. Despite the evenness, especially in hard news some extreme outliers are observable, most prominently in Australia, and the content of these particular articles exemplifies some of the concrete functions passive voice can take in news. The most common themes in this domain are crimes and accidents, as was indicated by the keyword analysis; when dealing with these topics, some factors, like identities of criminals, can be unknown, which then makes the passive voice the most reasonable choice for a report, as example [25] shows. In other cases, the entity functioning as the subject in an active construction is obvious or simply not important, and the passive is used to shift the focus to another party, as in example [26]:

[25] North Kuta Police chief Aldi Alfa said Mr Gill **was attacked** by two people attempting to rob the villa. (Lamb 2013, *The Sydney Morning Herald*; emphasis added)

[26] The victim **was rushed** to the Royal Brisbane and Women's Hospital [...] (Bochenski 2013, *The Sydney Morning Herald*; emphasis added)

Yet, even if for instance the identity of an attacker is known, structuring the sentence in passive voice creates a different emphasis in a story as it foregrounds one news actor (in this case the victim) while omitting or backgrounding another (the attacker in example [25], the ambulance in example [26]) (Fowler 1991, 77-78). Van Dijk

(1998, 32-33; emphasis in original) describes a pattern with regard to this saying that "OUR people tend to appear primarily as actors when the acts are good, and THEIR people when the acts are bad, and vice versa: THEIR people will appear less as actors of good actions than do OUR people". A study presented by Bohner (2001) furthermore found that especially in crime news, the passive is frequently used to assign responsibility to the victim by omitting the assailant as an actor. An ideological construction of the roles of good and bad can thus be strengthened by a journalist's choice of voice.

Van Dijk's approach, originating from Critical Discourse Analysis, concentrates more on news with ideological dimensions, and in the present dataset respective examples can be found primarily in political news featuring violence or crime:

> [27] Egypt's new military-backed government said its soldiers acted in self-defense during the altercation, in which soldiers gunned down at least 50 Islamists. One Egyptian soldier **was killed**. (Bradley 2013, *The Wall Street Journal Asia*; emphasis added)

Here, the roles of good and bad are created by the verbs *gun down* and *killed*, and while the former is clearly attributed to an actor, the second one is phrased in passive voice and thus requires more effort from the reader to identify the culprit. It is obvious which of the two involved parties the audience is supposed to feel close to, and towards which one the author increases the distance.

It can thus be seen that passive voice is a powerful tool to identify in- and outgroups within a story. Using the passive always implies a higher distance towards the backgrounded or omitted party, while active voice is more straightforward and brings the reader closer to the respective news actors, marking them as active members of society. On the other hand, as the passive voice backgrounds the acting party, any evaluative items are also less effective to characterise them, which renders the style less involved and more neutral, thereby increasing the distance between author and reader as well.

In this sense, the results show that hard news and political news work with a higher distance than the other domains, which makes sense as they also displayed high tendencies towards factual information and lower values for evaluative features. Also in accordance with previous results, lifestyle and sports news contain most active constructions and thus underline their interpersonal focus. Among the varieties, Kenyan and Hong Kong news writing once again stands out as being more distant, while

US news appears most directly oriented towards their readerships. As was already the case for boosters and minimisers, there is no significant difference concerning voice between more quality and more popular papers neither with regard to overall numbers nor within any domain except lifestyle; here, quality papers contain significantly fewer passive constructions than expected at χ^2=105.06 (df=1, p<0.000), thus indicating that in these news items, quality papers tend towards a lower distance more clearly than tabloids.

Nouns and nominalisations

Similar to passive constructions, a high amount of nouns in a text reflects a more distant style as the emphasis is shifted away from the action and respective agent. However, in contrast to voice, where the author has the choice between active and passive, nouns are of course not entirely avoidable in a text and are carriers of information, which makes them essential in news. While the overall frequencies of nouns are interesting and will be taken into account, they are more relevant for the subdimension of medium (section 5.3.1); more telling at this point are instances of the linguistic transformation of nominalisation, as they reflect concrete cases in which the author decided for a noun and against a verbal construction.

For this investigation, instances of nominalisation were queried as lemmata with a nominal part of speech-tag (NN for singular noun and NNS for plural noun) and an ending of *-tion*, *-ness*, *-ity*, *-ism*, *-ment* or *-ing*. The item *thing* was excluded from the query. Proper names (part of speech-tags NP and NPS for singular and plural) were not taken into account as names are given entities and are not subject to the authors' decision-making process in this sense. Table 5.19 lists the raw frequencies of nominalisations as well as all nouns (the latter including the amount of the former), their respective relative frequencies in relation to all word tokens per category, and the overall values per domain and variety in the last column and row.

Table 5.19: Frequencies of nouns and nominalisations

		AUS	HK	KEY	UK	USA	Σ *domains*	*range*
ECO	Nomin.	1,992	2,536	2,558	1,905	1,959	10,950	653
	%	2.8	3.09	3.89	2.25	2.3	2.82	1.64
	Nouns	17,164	20,760	16,199	20,145	20,767	95,035	4,568
	%	24.15	25.33	24.66	23.83	24.36	24.46	1.5

HN	Nomin.	857	1,590	1,498	1,262	1,879	7,086	1,022
	%	1.56	2.45	2.57	1.53	2.2	2.05	1.04
	Nouns	12,263	15,540	13,486	17,656	19,228	78,173	6,965
	%	22.39	23.93	23.11	21.38	22.54	22.6	2.55
LIFE	Nomin.	1,787	2,359	2,890	1,694	1,910	10,640	1,196
	%	1.76	2.01	2.29	1.48	1.62	1.84	0.81
	Nouns	22,555	28,080	27,872	24,405	26,689	129,601	5,525
	%	22.25	23.88	22.05	21.38	22.69	22.46	2.5
POL	Nomin.	2,614	3,284	2,648	2,264	2,435	13,245	1,020
	%	3.41	3.99	3.53	2.56	2.73	3.22	1.43
	Nouns	16,635	18,621	15,156	18,541	19,141	88,094	3,985
	%	21.71	22.63	20.21	20.94	21.49	21.41	2.42
SPO	Nomin.	953	1,560	912	1,184	1,059	5,668	648
	%	1.19	1.81	1.41	1.19	1.08	1.32	0.73
	Nouns	14,455	17,507	12,537	17,002	17,369	78,870	4,970
	%	18.03	20.27	19.37	17.02	17.78	18.39	3.25
Σ	Nomin.	8,203	11,329	10,506	8,309	9,242	47,589	
variety	%	2.14	2.62	2.69	1.77	1.95	2.21	
	Nouns	83,072	100,508	85,250	97,749	103,194	469,773	
	%	21.36	23.2	21.85	20.81	21.73	21.83	
range	Nomin.	1,757	1,724	1,978	1,080	1,376		2,427
	%	2.22	2.18	2.48	1.37	1.65		2.91
	Nouns	10,292	12,540	15,335	7,403	9,320		15,817
	%	6.12	5.06	5.29	6.81	6.58		8.31

Again, the differences that can be observed are bigger between the domains than between the varieties. News from economy and politics produces the highest amounts of nominalisations, whereas in terms of all nouns, politics is more moderate and only economy is clearly set apart. Sports news on the other hand contains fewest nominalised forms as well as fewest nouns in general. The variation between the domains is almost equally significant in terms of nouns (χ^2=4844.17, df=4, p<0.000) and nominalisations (χ^2=4566.41, df=4, p<0.000), and in both cases, the extremes of economy and sports news are the strongest factors.

Looking at the varieties, news from the UK shows the lowest frequencies of nominalisations in all domains except sports. The L2 varieties Kenya and Hong Kong again demonstrate a higher distance by including most nominalised items; the former is outstanding in news from economy, hard news and lifestyle, whereas the latter dominates in sports news and politics. In contrast to the domains, the varieties on average differ less in their frequencies of all nouns; the use of nominalisations thus appears to have an influence on the overall numbers in the sense that more nominalised forms imply fewer other nouns and vice versa, thereby evening out the figures across the varieties. The differences in both features are still significant though, at χ^2=775.54 (df=4, p<0.000) for nouns and χ^2=1334.43 (df=4, p<0.000) for nominalisations.

Since the variation is expressed so strongly on the level of domains, the varieties do not display any particular clusters when regarded on the basis of their averages. The domains on the other hand cluster more clearly, as a distribution of all texts shows. Figure 5.25 displays the individual texts along the lines of nominalisations and other nouns, both in relation to their total word tokens, sorted by their domains.

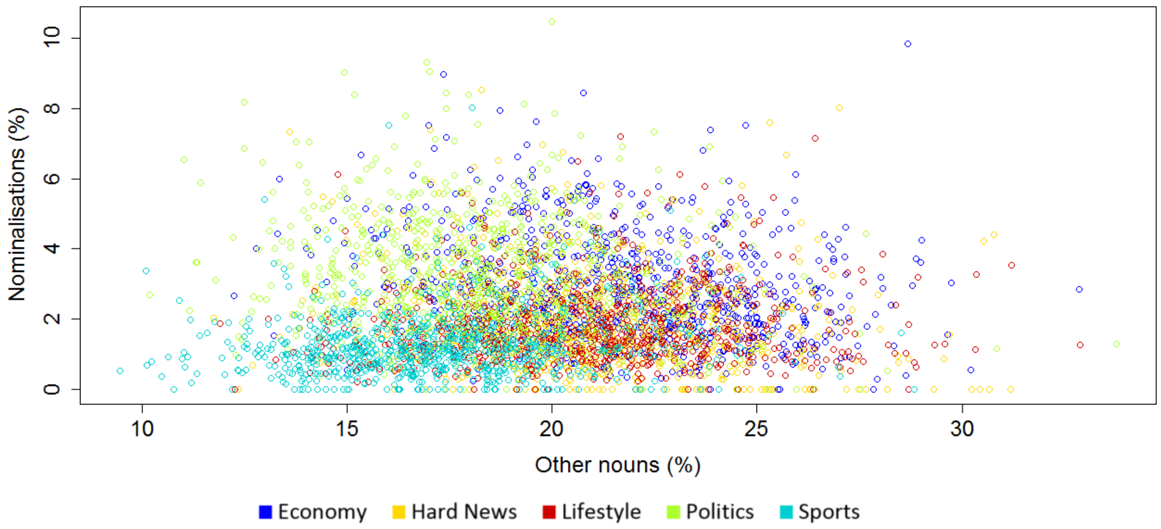


Figure 5.25: Nouns versus nominalisations

From this perspective, it becomes more apparent that with regard to other (non-nominalised) noun forms, the most widespread domain is hard news, whereas politics produces the broadest range on the dimension of nominalisations. Parallel to this, political news does not stray far into higher numbers of other nouns, while economy makes frequent use of both noun-forms. The high cumulative noun frequencies in these two domains confirm the trends that became visible with regard to the field of discourse, namely that these news stories focus on facts and therefore contain many lexical items in contrast to function words and tend towards a high degree of informativeness. The frequent use of nominalisations underlines this further and suggests that the authors very consciously decrease the emphasis their stories put on the actions and actors and channel the readers' attention to specific aspects:

> To understand this, reflect on how much information goes unexpressed in a derived nominal, compared with a full clause: compare, for example, 'allegations' with the fully spelt-out proposition 'X has alleged against Y that Y did A and that Y did B [etc.]'. Deleted in the nominal form are the participants (who did what to whom?), any indication of time – because there is not verb to be tensed – and any indication of modality – the writer's views as the truth or the desirability of the proposition. (Fowler 1991, 80)

In this context, the high numbers of nominalisations in Hong Kong and especially Kenyan news can be interpreted as politically motivated, as they decrease the evaluative content of the text and suppress the author's opinion, which appears logical in countries with a low press freedom. Nevertheless, it should be kept in mind that the results here were not filtered by whether or not they appear in direct quotes – the omission of information implied in a nominalisation can therefore also originate from other actors than the journalist. To trace these instances a qualitative analysis is required which looks not only at the nominalised expression but also at if and how the journalist fills the gaps created by the syntactic choice at another point in the article.

A high distance and neutral style is less critical in the domains of sports and lifestyle, as the topics treated here are less ideologically loaded. Table 5.19 already showed that sports news contains fewest nouns on average, which is confirmed by the distributions of the individual texts. Lifestyle articles on the other hand include relatively many non-nominalised nouns, but only few nominalisations. While sports therefore appears to put its emphasis on actions in the forms of verbs and adjectives, lifestyle news is more similar to the other domains in terms of overall frequencies, but tend to use nouns less frequently as alternatives for other syntactic structures, thereby keeping the distance between readers and authors at a lower level.

5.2.3 Summary

All in all, twelve linguistic markers were analysed in this section, four as operationalisations of social role relationship and thirteen for social distance, with modality occurring in both subdimensions. Summarising the aspects of the tenor of discourse it can be said that although results are very diverse for the individual linguistic features, some tendencies showed repeatedly and can be assumed to be characteristic for the respective domains and varieties.

Looking at the domains, most outstanding in this regard are the news from lifestyle sections. The items published under this heading proved to contain the highest amounts of interrogatives, shorts words and a low specialisation, making them most accessible for a broad audience and constructing small gaps in expertise and authority. With regard to social distance, this orientation towards the audience was confirmed by a high degree of evaluation in the language, frequent use of 1st and 2nd person pronouns and contractions as well as few instances of nominalisations and passive voice. Furthermore, the analysis of place names indicated that lifestyle articles aim at local audiences, hardly ever referring to places outside their own country.

While in terms of social distance sports news proved to be very close to lifestyle articles, these items showed a much higher specialisation in their vocabulary. In combination with a low average regarding word length, it can be concluded that sports news aims at constructing closeness between author and reader, but require an interest in the subject that is not just general, but strong enough to include expertise. The potential audience is therefore not as broad as for lifestyle news, although the stories are told in a similarly informal style.

Hard news was at the other end of the scale for distance in many aspects. This domain produced the lowest values for modal verbs in all regions and contained fewest boosters and minimisers and evaluative adjectives as well, thereby showing a clear orientation towards neutral style and factual information unobscured by speculation or ambiguity. It also stood out in terms of high frequencies of passive constructions however, which, although signalling a higher distance in general, serve to highlight aspects and actors and can therefore be regarded as evaluative tools as well.

The other domains, economy and politics, clustered together more closely in overall terms, yet especially political demonstrated particularities in some aspects and stood out with fairly mixed tendencies towards evaluative language, particularly with their high frequencies of modal verbs and titles and forms of address. These results were complemented by the highest numbers of nominalisations. This constellation suggests that while these news items create a formal impression and atmosphere, they also contain high degrees of vagueness and avoidance of clear statements, which function as indirect markers of evaluation.

Regarding the varieties, a rough distinction could be seen between the UK and Australia on the one and Hong Kong and Kenya on the other hand, although the latter two varied in several features. In terms of social role relationship news from Australia and the UK displayed the highest amounts of deviation from declarative mood as well as most modals, and the UK demonstrated a particularly low degree of specialisation as well. These indicators for a low discrepancy in authority and expertise were underlined by frequent use of boosters and minimisers and high numbers of evaluative adjectives and adverbs, suggesting a trend towards evaluative language. The use of 1st and 2nd person pronouns confirmed this further, revealing high values for the two regions.

Hong Kong and Kenya often appeared as opposites to this involved style; the former proved to be most rigorous in use of declarative mood and contained fewest modals and 1st and 2nd person pronouns, whereas it was more moderate in its isolation from the native varieties in terms of passive voice, boosters and minimisers and contractions. In these aspects, Kenya produced the most extreme values, albeit followed directly by Hong Kong in most cases. In terms of nominalisations and mean word length, the two regions were similarly set apart from the rest.

Least clear in their tendencies are news items from the USA. While they displayed the highest specialisation on average of all regions, they proved more informal with high frequencies of contractions and very few nominalisations, thereby creating a closer relationship between reader and author. The variety furthermore stood out in the analysis of place references, where it was the only region not to put an emphasis on geographical nearness but relating to European countries much more frequently than to its immediate neighbouring nations. Similar to Australia however, it also showed a strong focus on regional and local events, which is to be attributed to the size of the two countries.

The aspect of tenor of discourse therefore resulted in some interesting groupings regarding both varieties and domains. As became apparent in some instances during the analyses however, the functions of the individual features vary and are not always consistent throughout all their occurrences. Although they were interpreted here in the sense of what they have been found to represent most strongly in previous studies, this diversity means that some fine-grained implications get lost in a quantitative study. In the analysis of the last parameter of the situational context, the mode of discourse, some linguistic markers are taken up again and investigated from a different angle. The features will therefore not be put into relation at this point, but in the context of a bigger picture at the end of the analysis.

5.3 Mode of Discourse

The last parameter to define the situational context of a discourse is the mode, or “the particular part that the language is playing in the interactive process” (Halliday and

Hasan 1985, 24). As was seen in section 3.2.3, it is usually divided into the subdimensions of medium, channel and language role, of which primarily the first one is relevant for the investigation at hand.

5.3.1 Medium

The aspect of medium reflects the style of the language in use in terms of its properties of typically spoken or typically written discourse. The texts analysed in this study are all presented using the graphic channel, which means that they appear in writing and are to be read, not listened to. Nevertheless, they can contain linguistic characteristics which are usually connected to spoken language, thereby creating a different atmosphere in the news item and constructing a closer or more familiar relationship to the reader. Like the other subdimensions, medium should not be interpreted in isolation but is closely linked especially to the parameter of tenor and its sublevels of social role relationship and social distance, which is reflected in the operationalisation as well. Medium will be analysed via the features of pronouns, conjunctions, contractions, nouns and nominalisations and the distribution of individual word classes, many of which already occurred in the previous sections. At this point, the results will be viewed from a different angle and interpreted with relation to spoken and written style; tables and figures that were shown before will not be repeated, but will be referred to where necessary or complemented by other representations.

Pronominal use

Use of personal pronouns is the first indicator to be analysed for the subdimension of medium. Separate aspects of this feature have already been described in sections 5.1.2 (goal orientation) and 5.2.2 (social distance), but for the present parameter, not individual pronouns but their collective frequencies are of interest. The raw and relative frequencies of personal pronouns for every category as well as for the whole domains and varieties are summarised in Table 5.20.

Table 5.20: Frequencies of personal pronouns

	AUS	HK	KEY	UK	USA	Σ *% of domain*	*range*
Economy	2,440	1,943	1,816	3,056	2,664	11,919	1,240
	3.43%	2.37%	2.76%	3.61%	3.13%	3.07%	1.24
Hard News	2,832	2,062	2,754	5,399	4,223	17,270	3,337
	5.17%	3.17%	4.71%	6.53%	4.95%	4.99%	3.36
Lifestyle	6,837	5,786	8,936	8,471	6,727	36,757	3,150
	6.74%	4.92%	7.06%	7.42%	5.72%	6.37%	2.5
Politics	3,378	2,724	3,128	4,564	3,880	17,674	1,840
	4.41%	3.31%	4.17%	5.15%	4.36%	4.29%	1.84
Sports	5,306	4,270	3,090	7,238	6,342	23,246	4,148
	6.61%	4.94%	4.77%	7.25%	6.49%	6.12%	2.48
Σ	20,793	16,785	19,724	28,728	23,836	109,866	
% of variety	5.41%	3.87%	5.05%	6.11%	5.02%	5.12%	
range	4,397	3,843	7,120	5,415	4,063		7,120
	3.31	2.57	4.3	3.81	3.36		5.05

Among the domains, lifestyle and sports appear close together at the high end of the scale, with overall percentages of 6.37 and 6.12. In direct contrast to this is news from economy, which produces the lowest values in every domain, while hard news and politics are again more similar to each other. The differences in the direction of domains are highly significant at χ^2=6709.18 (df=4, p<0.000), of which the low numbers in economy make up a striking 3159.04.

Along the varieties, differences regarding pronominal use are most clearly expressed in the case of Hong Kong. While the other varieties all display values within a range of 1.09 percentage points, the discrepancy of Hong Kong news towards the second lowest variety, the USA, is 1.15, and towards the UK with the highest value 2.24 percentage points. Overall, the variation along this line is highly significant as well at χ^2=2430.05 (df=4, p<0.000), with Hong Kong and the UK contributing most to this value with 1284.71 and 940.22 respectively.

These tendencies are little surprising when put into context with the previous analyses. Both lifestyle and sports news demonstrated trends towards a low distance and a more casual style, which is underlined by a rather spoken orientation of the medium, whereas economic news tended towards a higher degree of distance and more formality, which in turn is emphasised by a style more common of written language. The same holds true for the variety of Hong Kong, in contrast to which the news from Kenya shows a more diverse picture by demonstrating a high distance with regard to evaluative language, contractions and nominalisations, but producing average values of pronominal use. The UK and Australia confirm their previously suggested tendencies towards a more direct link between author and reader by clearly displaying the strongest leaning towards the spoken medium.

Sentence-initial conjunctions

In the case of conjunctions, an assignment to spoken or written discourse is less straightforward, as different findings have been reported on their frequencies. Coordinating conjunctions, which in this corpus are marked with the part of speech-tag CC, have been found to be more common in spoken language by Neumann and Fest (2016, 210-11), but more characteristic of written discourse by Biber et al. (1999, 81). In both cases, the differences are relatively small, whereas very clear tendencies are observable for coordinators in sentence- or turn-initial position. Biber et al.'s (ibid., 84) findings suggest that in spoken language, coordinators occur at the beginning of a sentence or turn more than twice as often as in forms of written discourse. Table 5.21 lists the relative frequencies of coordinating conjunctions and of those occurring in sentence-initial position, both in relation to the word tokens of the categories, and the values for the whole domains and varieties.

Table 5.21: Relative frequencies of coordinators

		AUS	HK	KEY	UK	USA	*domain*	*range*
ECO	Coord.	2.7	2.6	2.88	2.61	2.56	2.66	0.32
	Sent.-init.	0.17	0.13	0.05	0.27	0.2	0.17	0.22
HN	Coord.	3.11	2.85	2.76	3.03	2.84	2.92	0.35
	Sent.-init.	0.07	0.1	0.06	0.17	0.1	0.11	0.11
LIFE	Coord.	3.71	3.6	3.42	3.53	3.45	3.54	0.29
	Sent.-init.	0.2	0.2	0.19	0.33	0.22	0.23	0.14
POL	Coord.	2.75	2.69	2.71	2.87	2.6	2.72	0.27
	Sent.-init.	0.18	0.12	0.1	0.21	0.22	0.17	0.12
SPO	Coord.	3.09	2.95	2.73	3.15	2.95	2.99	0.42
	Sent.-init.	0.17	0.16	0.12	0.31	0.29	0.22	0.19
variety	Coord.	3.12	3	2.98	3.07	2.92	3.01	
	Sent.-init.	0.17	0.15	0.12	0.27	0.21	0.19	
range	Coord.	1.01	1	0.71	0.92	0.89		1.15
	Sent.-init.	0.13	0.1	0.14	0.16	0.19		0.28

When looking at all coordinators, the overwhelmingly dominant item is *and*, which contributes 50,116 hits, making up 77.25% of all coordinators. It is followed by *but* (8,953 hits, 13.8%) and *or* (4,623 hits, 7.13%). The variation in terms of use of coordinators among the domains is significant at χ^2=465.74 (df=4, p<0.000), while the differences between the varieties are mirrored in a χ^2-value of 123.46 (df=4, p<0.000).

Regarding the varieties, the distribution displayed here is in line with what has been found for previous features. News from Australia and the UK, which produce the highest values here, also contained most evaluative adverbs and adjectives and most personal pronouns, in general as well as in terms of 1st and 2nd person items. These linguistic markers all pointed towards a less distant and more informal style, which suggests that contractions reflect spoken rather than written language. However, these varieties displayed the highest frequencies of titles and forms of address as well. Among the domains, the results are even more difficult to interpret in terms of written and spoken medium, as lifestyle and sports have shown trends towards informality and closeness before, which would indicate their high amounts of conjunctions as a representation of written medium, yet hard news has so far often appeared at the exact opposite end, while with regard to coordinators, it is similar to lifestyle and sports news.

Looking at the amount of sentence-initial coordinators, the numbers are more in line with previous results for the domains. Sports and lifestyle items are set apart with high values, whereas hard news produces the lowest and politics and economy cluster together with a clear tendency towards the lower end. Among the varieties, the UK confirms its previous tendencies towards a closer relationship between author and reader by demonstrating a more spoken style, whereas Kenya and Hong Kong underline their emphasis of distance and more standardised newspaper language by displaying lower values for sentence-initial coordinators and thereby orienting themselves towards the written end of medium.

Contractions

The linguistic feature of contractions has already been discussed in context of social distance (section 5.2.2), and its functions in this subdimension and medium are closely related. The use of contractions is very typical for spoken discourse and is used much less frequently in written language (Biber et al. 1999, 1128-32). Contracting word forms therefore signals a more direct style which has an influence on the distance between the discourse participants.

As was seen in Table 5.17 and Figure 5.23, especially news from Kenya proved outstanding in this regard with much lower amounts of contracted forms than statistically expected (0.17% of all word tokens). This trend was visible in all domains, whereas the USA produced the highest values in all areas and contractions made up

1.05% of its total word count. Hong Kong, too, was below average, although less clearly than Kenya, and Australia and the UK were statistically higher. It can thus be said that the L2 varieties tend towards written medium in contrast to the native varieties, which contain more elements typical for spoken discourse.

The results were similarly clear for the different domains. Lifestyle and sports news contained most contractions with 0.98% and 1.18% respectively, whereas hard news (0.4%), economy (0.35%) and politics (0.43%) contained fewer hits. This underlines the trend of lifestyle and sports news towards the spoken medium as well as of the other domains to tend towards characteristics of written discourse, which was already suggested by the numbers of pronouns.

Nouns and nominalisations

The linguistic marker of the use of nouns and nominalisations, too, has been discussed before (see section 5.2.2). While it was interpreted along the lines of its function to omit information or shift the emphasis of a statement, in the context of medium it is a logical complement to the frequency of personal pronouns. Pronouns, as was seen at the beginning of this section, are more common in spoken than in written discourse, which in turn implies that in written language, nouns and names are preferred. In contrast to the analysis for social distance, where names and nouns were kept separate, they can therefore be investigated in combination at this point. Table 5.22 lists the relative frequencies of nouns (part of speech-tags NN and NNS) and proper names (part of speech-tags NP and NPS), their cumulative percentages and, again as a subcategory included also in the former, the frequencies of nominalisations.

Table 5.22: Relative frequencies of nouns, nominalisations and proper names

		AUS	HK	KEY	UK	USA	*domain*	*range*
ECO	Prop. N.	9.63	10.9	10.53	8.48	10.28	9.94	2.42
	Nouns	24.15	25.33	24.66	23.83	24.36	24.46	1.5
	∑	33.78	36.23	35.19	32.31	34.64	34.4	3.92
	(Nomin.)	2.8	3.09	3.89	2.25	2.3	2.82	1.64
HN	Prop. N.	10.41	10.44	10.55	8.95	11.47	10.35	2.52
	Nouns	22.39	23.93	23.11	21.38	22.54	22.6	2.55
	∑	32.8	34.37	33.66	30.33	34.01	32.95	4.04
	(Nomin.)	1.56	2.45	2.57	1.53	2.2	2.05	1.04
LIFE	Prop. N.	6.42	8.77	6.52	5.68	8.98	7.3	3.3
	Nouns	22.25	23.88	22.05	21.38	22.69	22.46	2.5
	∑	28.67	32.65	28.57	27.06	31.67	29.76	5.59
	(Nomin.)	1.76	2.01	2.29	1.48	1.62	1.84	0.81

POL	Prop. N.	12.41	12.3	14.54	10.64	12.49	12.41	3.9
	Nouns	21.71	22.63	20.21	20.94	21.49	21.41	2.42
	∑	34.12	34.93	34.75	31.58	33.98	33.82	3.35
	(Nomin.)	3.41	3.99	3.53	2.56	2.73	3.22	1.43
SPO	Prop. N.	11.77	11.78	16.89	11.13	11.68	12.37	5.76
	Nouns	18.03	20.27	19.37	17.02	17.78	18.39	3.25
	∑	29.8	32.05	36.26	28.15	29.46	30.76	8.11
	(Nomin.)	1.19	1.81	1.41	1.19	1.08	1.32	0.73
variety	Prop. N.	9.9	10.69	11.06	8.85	10.87	10.26	
	Nouns	21.36	23.2	21.85	20.81	21.73	21.83	
	∑	31.26	33.89	32.91	29.66	32.6	32.09	
	(Nomin.)	2.14	2.62	2.69	1.77	1.95	2.21	
range	Prop. N.	5.99	3.53	10.37	5.45	3.51		11.21
	Nouns	6.12	5.06	5.29	6.81	6.58		8.31
	∑	5.45	4.18	7.69	5.25	5.18		9.2
	(Nomin.)	2.22	2.18	2.48	1.37	1.65		2.91

Looking at the varieties, it can be seen that the UK contains the lowest frequencies of nouns and names as well as nominalisations. This is compatible with the high results for pronominal use and confirms the variety's tendency towards the spoken medium. Australia, too, shows low cumulative values of nouns and names, but more nouns are realised in the form of a nominalisation. News from the USA is the opposite of this; here, nouns and names are in general more frequent, yet nominalisations are comparatively rare. Kenya and especially Hong Kong contain more nouns and names, and the L2 varieties also display the highest amounts of nominalised structures, thereby again confirming their previously suggested orientation towards a more written style.

Among the domains, lifestyle and sports news, too, confirm previous results by displaying low numbers of nouns, names and nominalisations. A difference between the two domains can be seen with regard to the distribution; sports news contains the second highest amount of names of all domains, in contrast to lifestyle, which contains fewest. In contrast to this, nouns and nominalisations are represented more strongly in lifestyle, which reflects the topics of the respective items. In sports, individual athletes, coaches, teams, organisations and places are the dominant actors in the stories, as was shown by the keyword lists. In lifestyle news on the other hand, the news value of eliteness, which puts the focus on the popularity and importance of the involved actors, is less relevant, and the analysis of place references showed that location is least important in this domain as well, which decreases the number of names to be mentioned. Despite this difference however, it can be said that both domains tend towards a more spoken style than the other news, which contain more nominal items as well as nominalisations.

Lexical Density

The last feature to be analysed in this section is that of lexical density. Having been a part of the subdimension of experiential domain (section 5.1.1), a lot has been said already about the differences between domains and varieties in this regard. The values which were represented in Figure 5.8 are summarised in Table 5.23.

Table 5.23: Lexical density

	AUS	HK	KEY	UK	USA	*domain*	*range*
Economy	60.1	62.64	61.06	60.16	61.37	61.27	2.54
Hard News	60.35	61.22	60.38	58.75	60.9	60.27	2.47
Lifestyle	58.98	60.57	57.55	57.58	60.09	58.94	3.02
Politics	61.03	61.71	60.34	59.34	61.22	60.72	2.37
Sports	58	59.14	60.9	56.62	57.26	58.18	4.28
variety	59.77	60.99	59.65	58.38	60.1	59.76	
range	3.03	3.5	3.51	3.54	4.11		6.02

As lexical density is a representation of the relation between content and function words, it includes the values of individual word classes that were discussed previously. A low lexical density is characteristic of a more spoken style (Neumann 2013, 76), which especially in the case of news from the UK and Hong Kong and from the domains of lifestyle and sports once more confirms what has been suggested by earlier results. The other varieties show less variation and produce similar values, as do the other domains.

A closer insight into the structures of the individual categories is given by a breakdown into the major function and content word classes. Figure 5.26 displays the distributions for the five domains, calculated across all varieties.

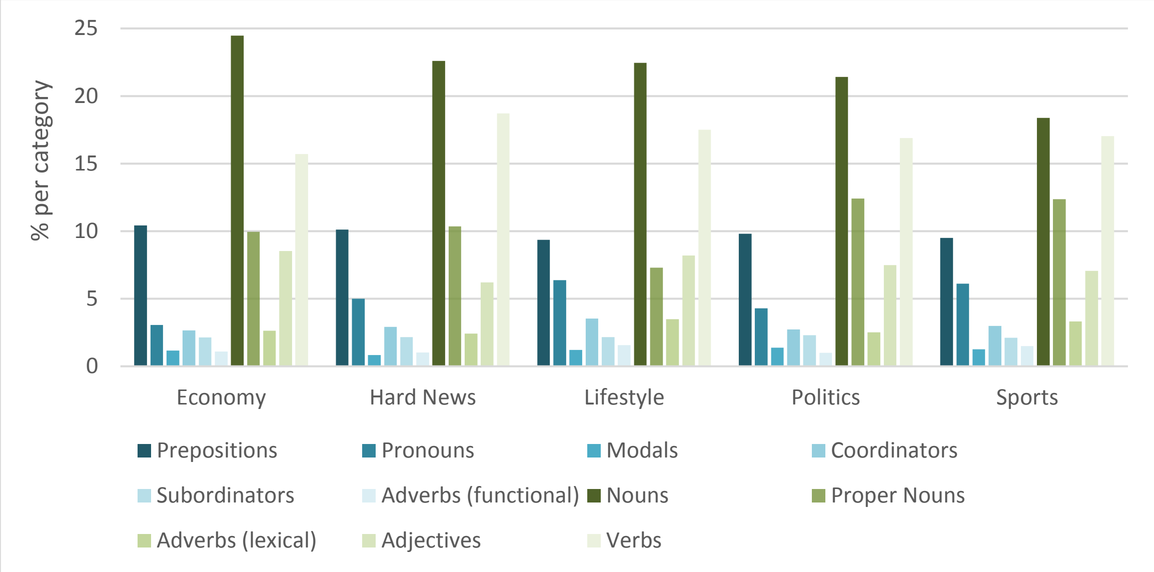


Figure 5.26: Distribution of word classes - Domains

In all domains, nouns and verbs are the dominant word classes, however, their difference is most strongly expressed in economic news and is almost neutralised in sports items. Proper names make up the third biggest content word class in all domains except lifestyle, where adjectives are more frequent.

Among the function words, similar universals can be observed with regard to prepositions and pronouns, which clearly dominate this field. Especially in lifestyle and sports news however, the difference between these two classes is less expressed, which is due to the high numbers of pronouns in these domains; the amount of prepositions on the other hand remains almost unchanged across all five areas. Subordinating and coordinating conjunctions are also very constant, but in lifestyle, sports and hard news articles, coordinators display a stronger dominance than in political and economic news.

These distributions confirm a lot of what has been indicated by the results for previous features. The high amount of verbs in sports news was already suggested by the keyword lists, and the low lexical density of lifestyle and sports, caused by fewer nouns and proper names as well as more pronouns, reflects the closer and less formal style of these domains. Economy on the other hand puts a clear focus on nouns and, as was seen in Table 5.22, on nominalisations as well, whereas it contains the fewest verbs. The high lexical density values for this domain thus originate from the dominance of nouns, whereas in hard news, verbs are a responsible factor and politics draws on a high amount of proper names. For all three, a tendency towards the written medium can be concluded.

Looking at the distributions for the varieties, across all domains, it can be seen that as was the case for many individual features, the differences in this direction are less expressed. Nevertheless, some reflections of what has been found before are observable (see Figure 5.27).

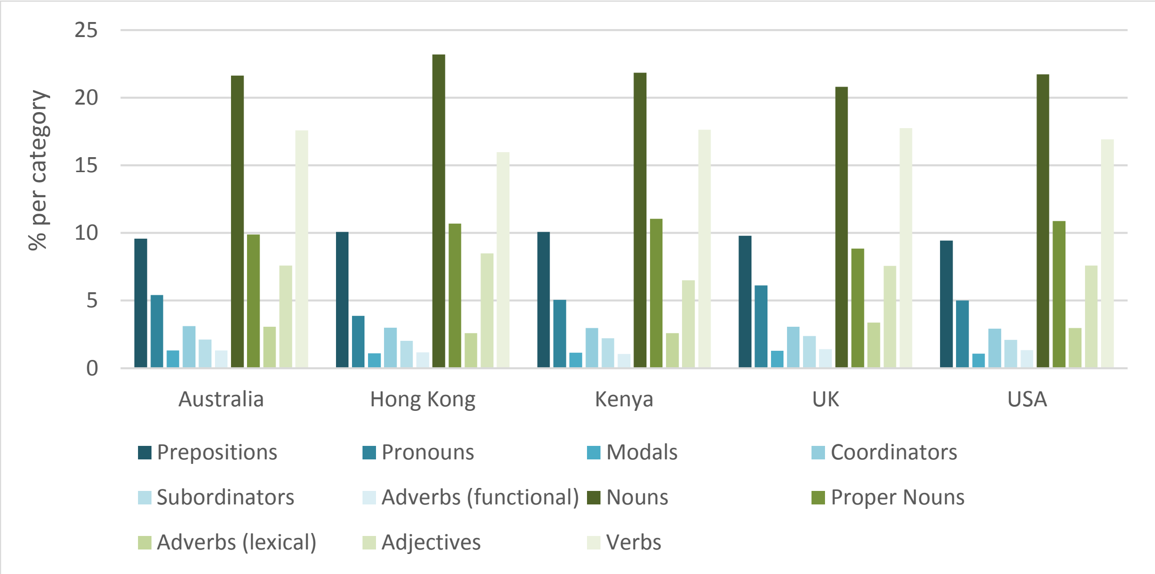


Figure 5.27: Distribution of word classes – Varieties

In this direction, too, nouns and verbs are the dominant word classes, although they again show differences in their relation, especially in the news from Hong Kong, which rely heavily on nouns, and the UK, which are more verb-based. The low lexical density which was observable for the UK can be traced back to a particularly low use of proper names and comparatively many pronouns. Among the function words, prepositions are again almost constant across all regions, but differences between coordinating and subordinating conjunctions are more clearly expressed in favour of the former than within the domains.

With regard to medium, the distributions confirm what was indicated previously, namely that the UK shows a much stronger tendency towards spoken style than the other varieties. The L2 varieties, but primarily news from Hong Kong, are at the other end of the scale within this dataset, demonstrating a clear orientation towards written style.

5.3.2 Summary

In contrast to the parameters of field and tenor, mode of discourse was represented in this study by only one subdimension, the medium. The operationalisation contained five features, pronominal use, sentence-initial conjunctions, contractions, nouns and nominalisations and lexical density. Since four of these had been analysed before in the context of another parameter, their results did not only contribute to an understanding of medium alone but could be put into relation to the previous insights and demonstrated close connections between the subdimensions and parameters.

In this sense, a number of assumptions resulting from the preceding analyses were confirmed by what was seen for medium. Again, more variance could be seen between the domains than between the varieties, and particularly sports and lifestyle news showed the clearest orientation towards the spoken medium in all five linguistic indicators. Among the varieties, although the values were generally closer together, some characteristics could be found which support the distinction between native and L2 varieties; news from the UK tended towards spoken style with many pronouns and sentence-initial conjunctions, a low lexical density and relatively few nouns and nominalisations, whereas Hong Kong proved to contain fewest personal pronouns and the highest lexical density on average. Regarding sentence-initial conjunctions and contractions Kenya and Hong Kong produced similarly low results and both demonstrated high frequencies of nouns and nominalisations.

Since the domains differed more strongly than the regions, the varieties show more internal variation than the individual domains and do therefore not display any obvious clusters in terms of their distribution. The domains on the other hand are reflected more clearly in this way, as Figure 5.28 shows.

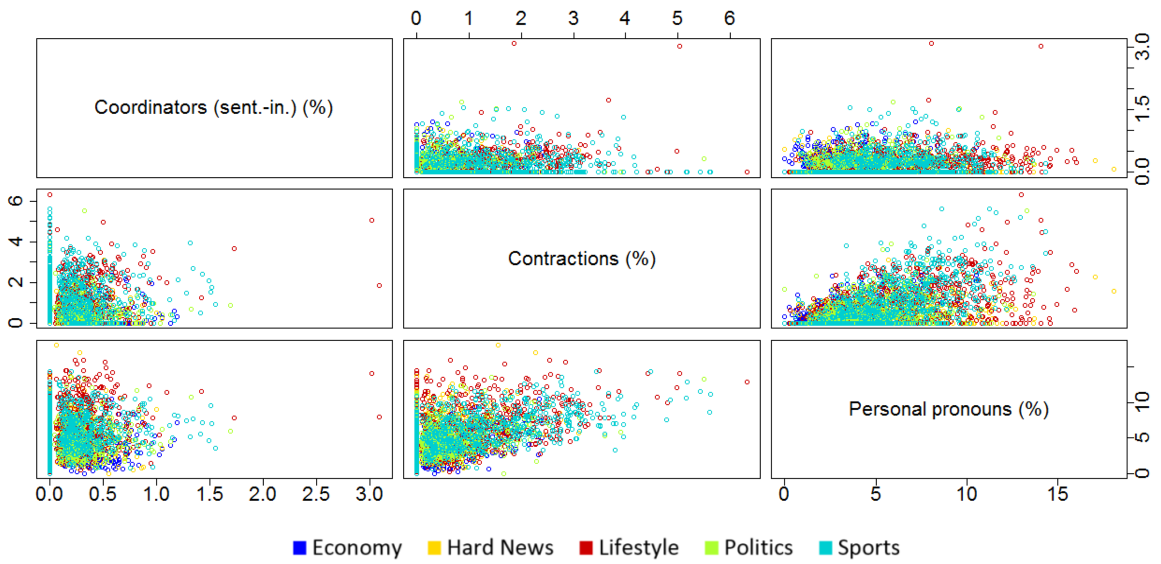


Figure 5.28: Correlations of indicators of medium

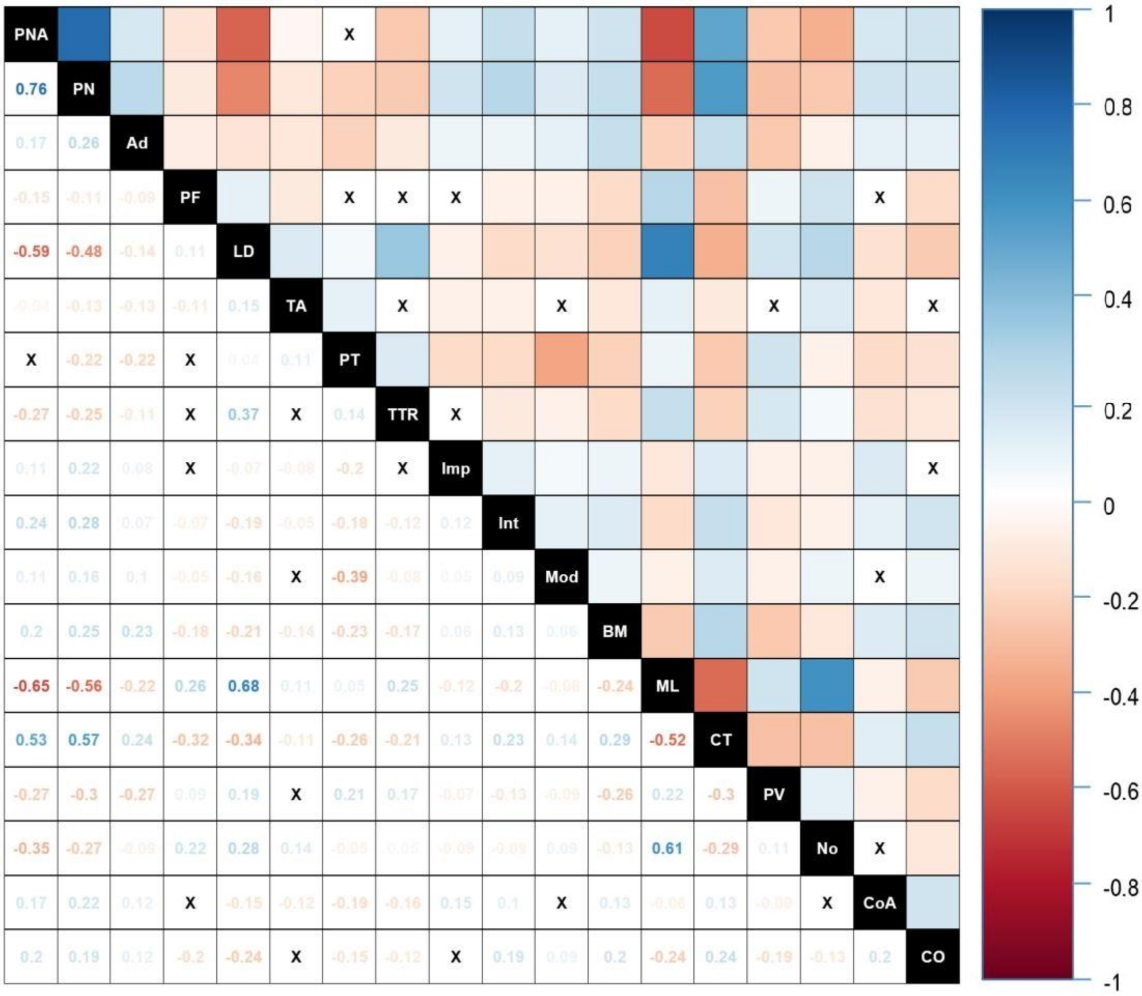
Due to the interdependency of some of the five features, the graph shown here was restricted to three variables. Although the majority of articles still forms one cloud, it becomes apparent in this representation that lifestyle and sports articles generally spread more widely and contain more extreme values of the individual features than the other domains. With regard to sentence-initial conjunctions, many texts do not display any instances, which creates a line along 0% for this feature, which can be observed, although less strong, for contractions as well. All three variables show correlations with each other, and particularly contractions and personal pronouns are dependent at r=0.53. As was said at the beginning of this section however, the features analysed for medium are reflective of other aspects of the situational context as well; before individual correlations will be interpreted, the following section will therefore summarise all linguistic aspects and control for dependencies.

5.4 Synopsis

The preceding sections analysed the newspaper articles in the corpus from three different perspectives, the field, tenor and mode of discourse, reflecting the topic, interpersonal relations and relevance and use of language. Together, these parameters define the situational context in which the language in question is used and allow differences between instances of language to be traced back to particular subdimensions and influences.

The analyses conducted and described in this chapter revealed a number of characteristics for the individual domains and varieties, but also about the methodology and the function of linguistic features and their assignment to the various subdimensions. The operationalisation described in chapter 3 drew on results and findings from previous studies with both linguistic and journalistic foci, and the analysis therefore inevitably also worked as a control mechanism for the suggested linguistic markers. While some features proved rather straightforward in their representativeness of situational aspects, in some cases the results, especially when complemented by qualitative spot tests, showed that features fulfilled various functions in the discourse which were not immediately visible in the quantitative results and did not add to the originally assumed effect, but implied some other intentions. The most obvious linguistic markers of the ones analysed here for which this discrepancy was found were interrogatives and imperatives; while the former varied mainly between being used as rhetorical questions and teasers and could be summarised as devices to lower the degree of authority constructed, the latter proved to be problematic as it was used both for giving advice, thereby emphasising expertise, and weakening a statement and asking for consent, thus lowering the level of authority and expertise. An unambiguous interpretation towards one way or the other can only be achieved by a manual analysis of every separate instance, whereas a quantitative analysis can only make very general assumptions.

Apart from such cases of contrary effects, many features were assigned to and analysed in the context of more than one subdimension. This does not imply that the functions of the respective linguistic markers were as diverse as those of for instance imperatives, but suggests that these features represent different aspects of the situational context. This multiple operationalisation of course means that the features within one subdimension cannot be expected to all correlate strongly – if a feature demonstrated perfect correlations with two or more different sets of features, the differences between the subdimensions would have to be questioned. A principal component analysis revealed no additional insights into the data, nevertheless, which features correlate and to what degree puts the overall results into perspective once more. Figure 5.29 shows these relations and indicates the direction and strength of a potential correlation, indicating matches that do not hold a significant correlation at a level of $p=0.01$ with a cross. It includes the majority of linguistic markers analysed in this study, excluding only those which could not be quantified suitably (e.g. lexical items and keywords). In addition, the Press Freedom Index, represented by the rank of every article's respective publication year, was taken into consideration as well.



PNA	Pronouns (all)	PT	Past tense	ML	Mean word length
PN	Pronouns (1st / 2nd p.)	TTR	Type-token-ratio	CT	Contractions
Ad	Eval. adv./adj.	Imp	Imperatives	PV	Passive voice
PF	Press Freedom Index	Int	Interrogatives	No	Nominalisations
LD	Lexical density	Mod	Modality	CoA	Coordinators (all)
TA	Titles / Forms of address	BM	Boosters / Minimisers	Co	Coordinators (sentence-initial)

Figure 5.29: Correlation patterns across all features

Since some of the features shown in this matrix are dependent on each other for structural or morphological reasons, these correlations are of course logical. Examples for such cases are the negative correlations between pronouns and lexical density as well as between mean word length and contractions, or the positive dependency of mean word length and nominalisations. The correlation between lexical density and mean word length, too, is not surprising, as content words, ergo nouns, names, adjectives, adverbs and verbs, are generally longer than the dominant function word classes of prepositions, determiners and pronouns.

Leaving these obvious correlations aside, some interesting connections are revealed which give an insight into the compatibility and dominant functions of individual features. The problematic case of imperatives for instance produces only low correlation values across the board, yet the positive ones are in relation to features which represent informal style and spoken medium, most prominently 1st and 2nd person pronouns, interrogatives and contractions, but also modality and boosters and minimisers. The strongest negative correlations on the other hand show in relation to past tense use, which is syntactically conditioned, and mean word length. Without looking at the individual instances of imperatives it can thus be said that the general trend of this linguistic feature points towards it being a marker of equality and informality, albeit only feebly.

A more clearly expressed tendency is observable for contractions. Having been found to be characteristic of spoken language as well as informal style in previous studies (Biber et al 1999, 166-67, 1128-32; Neumann 2013, 69), the use of contractions shows positive correlations with other markers of this nature, such as pronouns, interrogatives, sentence-initial coordinators and modality. Furthermore, it correlates with boosters and minimisers and evaluative adjectives and adverbs, thereby also including the dimension of evaluative language beyond modality.

Interestingly, contractions also demonstrate a negative correlation with the press freedom index value. Of all included variables, the rank of press freedom displays most insignificant connections, namely with use of past tense, type-token-ratio and imperatives. Many of the other correlations are comparatively weak, yet a tendency can be observed along the lines of the few stronger ones. Press freedom thus correlates positively with lexical density, mean word length, passive voice and nominalisations, which all reflect a high distance between the discourse participants and a low degree of evaluation. Especially passive voice and nominalisations are telling in this aspect as they were not considered direct markers of evaluative language from the start but were shown to have this dimension in the course of the analysis. Due to the omission of information which both of these features permit they appear as helpful linguistic devices for journalists facing strict censorship or political influence. In accordance with this, press freedom demonstrates a negative correlation with boosters and minimisers and evaluative adjectives and adverbs as well as with features indicating spoken style, which hints at a higher degree of standardisation.

The correlations depicted in Figure 5.29 draw on the values for every individual text and are thus more exact than the averages, yet they do not take into account any variation in terms of domain or variety. It was seen in the analyses that in general, the domains vary much more than the varieties and that especially lifestyle and sports news stood out as constructing a low level of distance and equal relations and tending most strongly towards the spoken medium. Among the varieties, this position was often held by news from the UK and Australia, whereas the L2 varieties demonstrated higher distances and more formal styles. Yet not only the absolute tendencies of these categories are of interest, but also their internal variation; for every feature that was analysed quantitatively, the difference between highest and lowest value was indicated in the respective tables, for every variety as well as every domain. These ranges reflect the level of standardisation in the respective news, and are much higher for the varieties than for the domains. Figure 5.30 shows the cumulative ranges for every region and topic.[28]

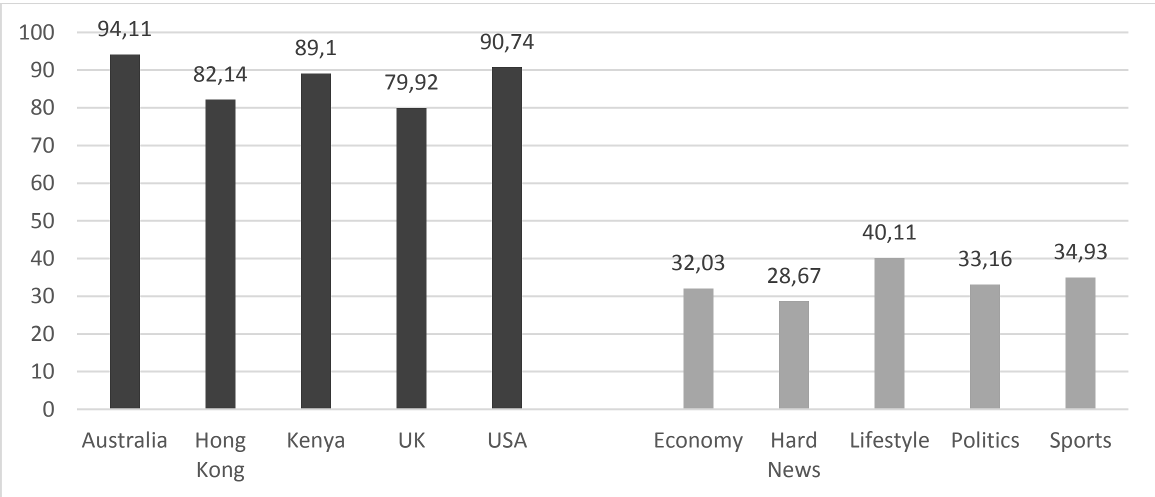


Figure 5.30: Cumulative internal variation

The differences between the domains for the individual features, which proved highly significant in the analyses, is of course reflected in the variety-internal ranges. In many cases, the lowest and highest values responsible for the ranges were contrib-

[28] A list of all ranges, averages and deviations for every feature for both varieties and domains can be found in appendix 4. Figure 5.30 focuses on the linguistic markers included in the operationalisation and does therefore not contain the values for raw article length. As Figure 5.7 showed, the differences between the domains and varieties in this regard are huge, which would have rendered the differences originating from the other features invisible in the graph at hand, which shows only whole domains and varieties. Instead, the corresponding values for raw article length will be included in the broader discussion in section 6.2, where the deviations of each category from its variety's average will be given (see Table 6.1).

uted by the outstanding domains of sports, lifestyle and hard news. The widest internal variation can be seen in Australia, whereas Hong Kong and the UK produce the lowest values. In the case of the latter, this in mainly due to little variation with regard to the use of past tense and type-token-ratio – in all other aspects, news from the UK is not outstanding, while Hong Kong shows low ranges across most features. For Kenya, the tendency towards the consistency in formality and distance, which was suggested by the analysis, is not reflected in the internal variation; in contrast to Hong Kong, the region displays an average amount of variation, although it is more similar to the low values of the UK and Hong Kong than to the high ones of Australia and the USA.

The domains, as was said above, display much lower ranges in general. Particularly hard news appears very standardised with little internal variation. The highest values are produced by lifestyle and sports articles, although the domain of sports clusters with economy and politics rather than with lifestyle news. The level of standardisation thus can be said to be lowest here, which confirms the extreme values this domain displayed for many features and the spread visualised for instance in the scatterplots in Figures 5.12 and 5.28.

The analysis of the parameters of field, tenor and mode in many ways produced compatible results for the domains and varieties, as was discussed in the individual sections and their summaries. The patterns that emerged and the differences that were found are not just representative of the individual subdimensions and parameters, but have to be regarded in the larger framework of media language and variational linguistics as well as register analysis. After the detailed and specific descriptions and interpretations offered in this chapter, the next chapter will therefore raise the discussion onto a more general level and connect the results to the theoretical concepts introduced earlier.

6 Discussion

It is the aim of this study to investigate differences in newspaper language along two dimensions, namely of regional and functional variation, and to use the insights gained from the analyses to draw conclusions about the developmental status of the individual varieties of English. For this, it is necessary to see which of the topical characteristics that were found in chapter 5 are common for the domains, and which ones are typical for a particular region. The following section will revisit the domains and discuss these issues and their implications, before they will then be put into the context of regional variation in the second section of this chapter.

6.1 Functional Variation – Characteristics of Newspaper Domains

The five domains included in this study were chosen not because of their content, but because of their universal presence in newspapers. Lifestyle, sports, economy and politics were separate categories in the analysed papers, while hard news was usually labelled just “news”. These domains therefore appear to be essential parts of press publications in all five regions, and could thus be analysed comparatively. Although they are not often told apart in journalism handbooks or guidelines, and writing advice is presented in general terms, some crucial differences could be found between the domains in terms of language use.

That the domains differ regarding the field of discourse, especially the experiential domain, is not very surprising. In each case, the keywords reflected the dominant topics in the domain and thus showed that the distinctions are quite concrete. Already for this very first linguistic feature, the domain of lifestyle was set apart by containing five pronouns among the top ten keywords, thereby indicating the focus on interpersonal matters rather than any concrete topics. This trend was confirmed throughout the analyses of social distance and social hierarchy as well as the high internal ranges for individual features and can be traced back to the wide scope and rather liberal structure of articles in this domain. For lifestyle news to be categorised as such, a human interest angle is desirable, which can be integrated into almost any subject. This is reflected further in the overwhelming use of present tense verbs and present references in these articles; when the focus is on interpersonal relations rather than events, the news value of recency is less important since the connection between reader and author or actor is happening at the time of reading, independent of the actual time of the event in question.

This dimension of interpersonality is constructed also by the use of a rather spoken style, which creates familiarity and also a less formal atmosphere. Especially the latter aspect makes it an expectable choice also for sports news; while economy, politics and hard news deal with more serious topics, sports news covers an area of life which is connected to free time and leisure for most people, and also holds less immediate relevance. It describes games, matches and teams, but even if the news item in question reports on a loss of one's favourite athlete or club, the direct impact on one's life is restricted to an emotional reaction. Other than for instance reports on an increase of taxes, election results, job cuts or strikes, sports news does not affect our lives permanently or in any material aspect and is therefore less serious in the sense that it cannot do much harm. This makes linguistic means such as irony or sarcasm possible which would be out of place in any of the more severe examples mentioned above, and also widens the structural options these news items can use.

Overall, sports news very often clustered with lifestyle and demonstrated a close relationship between author and reader and a very spoken style, which emphasises the distinction between formal and informal news domains. Furthermore, 1st and 2nd person pronouns were found to be used frequently, particularly 1st person singular ones, which indicates that either the author or, via a quote, a news actor is stating their personal opinion and thereby again construct a more personalised angle. Very fittingly, sports news also produced the lowest values for passive constructions and both nouns and nominalisations.

Despite this informal style and the high internal ranges indicating a low degree of standardisation, sports news items are more restricted in their topical scope than lifestyle news. As in the other domains, the keywords were very clear regarding the subject matter of this domain, but also showed that the focus is active by including a number of verbs rather than nouns. Additionally, sports news displayed the lowest type-token-ratio, which suggests that, although there are various sports, athletes and events that are covered, the choice is still more limited than in the other domains.

News from politics, economy and hard news stand in contrast to the interpersonal and casual style observable in articles dealing with lifestyle or sports. Especially hard news appears a very standardised domain, with the lowest internal variation and extreme values for many features. The keyness of the word *police* was striking in the analysis of the experiential domain and, as was already suggested in section 5.1.1, can be regarded to function as an external criterion which is reflected in the language directly. The temporal scope of this domain is oriented towards the past, and is combined with a particularly high distance between author and reader; there are hardly any imperatives or interrogatives in the hard news component, and evaluative language in the form of modals, boosters and minimisers and connoted adjectives and adverbs is rare, while titles and forms of address appear more frequently.

These linguistic indicators point towards a very factual tendency of the news in this domain, which appears logical considering that in most prescriptive journalism books, it is topics like the ones covered here, for example crime, natural disasters and accidents, which are given as standard examples for news stories and explanations for how to write them (e.g. Harcup 2005, 108-9; Keeble 2006, 94-108). They are, so to say, the stereotypical news, the ones to which principles like the six *W*-questions or the inverted triangle (Ungerer 1999, 24; Keeble 2006, 124-25) are most easily applied. At first sight, there is hardly any ideological loading included; for many stories, the parts of 'good' and 'bad' are easily assignable to victim and assailant and are defined by the mere cause of events, not by the journalist writing them up. However, as was seen in the analysis of voice, there are subtler means to include messages and opinions than using evaluative expressions outright. Hard news included by far the highest amount of passive constructions, an aspect in which they stray further off what is taught as journalistic language than any other domain despite their overall compliancy with these rules. This can be, and should be, attributed to a certain degree to formulations in which the actor omitted by the passive is not known or not relevant, as in phrases like "The victim was taken to hospital". Nevertheless, the points made by Fowler (1991), van Dijk (1998) and Bohner (2001) that passives are used to shift responsibility and blame are very relevant for hard news and have to be taken into account, meaning that this linguistic choice can also be used in an evaluating way, especially in articles which otherwise create an impression of objectivity and formality.

The remaining two domains, economy and politics, are the ones which are hardest to tell apart as their contents often overlap, and consequently, they have many linguistic characteristics in common. Probably because of their related topics, however, they also produce some very interesting oppositions. Their dominant subject matters were again very clearly reflected in their respective keyword lists, and concerning medium, both domains display a trend towards written style, using few contractions and sentence-initial coordinators and also including comparatively rare instances of pronominal use. Yet in this last respect, the news from these fields is set apart from the other domains by putting an emphasis on forms of the 1st person plural pronoun *we*, which is less expressed in the other parts.

In previous studies, the use of this particular pronoun has been analysed especially in connection with sports news, where it has been found to be used by the authors of news articles to refer to fans, including or excluding the respective athletes or teams depending on their success (Dunning 1999, 5; Fest 2011, 52-53). This competitive dimension is less relevant in news dealing with economy or politics, but a 'team-building' of author and readers against for instance negatively perceived political decisions or outgroups such as asylum seekers or refugees (cf. e.g. Gabrielatos and Baker 2008; KhosraviNik 2009, 2010a, 2010b; see also Elias 1978) would be a potential application for the 1st person plural pronoun. Spot tests for the current dataset reveal however that in most cases, these pronominal references are used in quotes, and here, they can have different referents, namely the readership, functioning as a direct address and an appeal to the community or nation, as in example [28], or the party or organisation in question, as in example [29]:

[28] Obama: "**We** are in a defining moment in **our** history. **Our** nation is involved in two wars and **we** are going through the worst financial crisis since the Great Depression." (Smith 2008, *The Sun*; emphasis added)

[29] But an ally of Mr Clegg took a dig at Mr Cameron, saying: "**We**, as a coalition government, will come to [the next] summit with some bold ideas about how **we** can increase growth, increase competitiveness and increase employment. And – yes – **we** will stay until the end." (Cooper 2011, *Daily Mail*; emphasis added)

While the first is an instance of a rhetorical device frequently found in political discourse when addressing problems or seeking solidarity (Pennycook 1994, 175-76; Bull and Fetzer 2006, 5), cases like example [28] demonstrate what a shift in speaker identity in other contexts can cause. In the given excerpt, it strengthens the speaker's position by implying that they are not alone, but have a group – or in this case, a party – behind them. When the news is bad or the speaker has to take responsibility for something, *we* can be used to shift the blame to several people or an abstract entity like a company, thus ensuring that the messenger in person does not become an immediate target (Pyykkö 2002, 246; cf. Mieroop 2005). The relatively frequent use of the 1st person plural pronoun in the domains of economy and politics can therefore be explained by these strategies which are applicable during election campaigns or company statements (cf. Fuoli 2013; Fuoli and Paradis 2014).

Despite these shared functions of this linguistic feature, the two domains differ greatly in other aspects, most prominently in the use of evaluative language and place references. Economic news, although creating the highest distance and highest degree of specialisation in terms of contractions, mean word length and lexical density, contain way more evaluative markers than political news. Especially regarding the feature of strongly connoted adverbs and adjectives, economy displays a difference of just 0.06 percentage points to the domain of lifestyle. News from the domain of politics on the other hand contains very few such items as well as boosters and minimisers, and instead make use of titles and forms of addresses most often.

This trend is confirmed by the domains' respective usages of nouns and nominalisations. Politics produces the highest values for nominalisations, but does not stand out in terms of overall nominal use, whereas economic news contains the highest amount of nouns in general, but slightly fewer nominalisations than political items. Since nominalisations were found to be a means of omitting information and avoiding commitment and responsibility, the vagueness constructed by using them is compatible with the more formal and distant style of political news, and sets the stage for speculations and promises while at the same time making the vocabulary appear more complex and technical. Articles from the domain of economy, too, include this feature to a higher degree than the other domains and falls only slightly short of its use in politics, which illustrates the relation between the two areas once more.

Furthermore, the domains display an interesting complementary distribution of place names: in all regions except the USA, the foci on international or national places in these two domains are oppositional, but show only a weak general trend; in Australia and Kenya, economy emphasises international, in Hong Kong and the UK national references, with the difference in the latter being insignificant. Only news from the USA concentrates on international aspects in both, but only these two, domains. In contrast to the indicators regarding distance and medium, which revealed generally coherent trends for the individual domains, this feature appears region-dependent and an explanation for this phenomenon is unlikely to be found in the nature of the domains themselves.

In general, contrary to the regional varieties, the different news domains are predefined by the newspapers' categorisation and assignment of journalists to specialised areas, which makes the domains a more artificial taxonomy than the regional varieties. It is therefore not surprising that the domains display a certain internal consistency and the overall use of some linguistic features is comparable in all regions. Particularly universal are the outstanding domains of sports, lifestyle and hard news, which are most clearly set apart regarding their respective topics as well as their descriptions in guideline books and manuals. Despite general trends however, crucial differences between the regions can be found in the degree of variation and the emphasis within individual parameters which reflect the status of the varieties and the foci of their respective societies.

6.2 Newspaper Language in the Context of Regional Variation

Regarding regional variation, as became apparent already in some aspects discussed in the previous section, the five varieties analysed in this study display differences that were expectable when looking at the regions' histories, geographical particularities and language policies, and that confirm the majority of assumptions formulated in section 2.4.

In general, a distinction between native and L2 varieties is observable, yet the individual features showed that it is particularly expressed in language markers related to objectivity and evaluation. In these aspects, the news from Kenya and Hong Kong proved to construct a higher distance between author and reader and draw less on evaluative indicators or personal statements, which could partly be related to the restricted press freedom in these regions. Despite this shared trend, a difference can be seen between the L2 varieties; in Hong Kong, interpersonal closeness was least expressed and the impression of objectivity emphasised most strongly, which confirms what has been found about the politically special situation of that region. Although officially a part of China, it still holds an autonomous status and, due to its strategic position as a pivotal point for trade and economy, is oriented towards international partners more than the other regions. This is reflected in the use of place references; while the news of all varieties expressed the news value of proximity by containing more national than international place names, in Hong Kong the international references made up 42.93% of all investigated. Only the UK, with 47.12% international ones, produced a higher value.

Nevertheless, the political system of China is influential on publications in Hong Kong (Lee and Lin 2006; Lee and Chan 2008). One aspect in which this shows is the general press freedom, which represents, among other factors, the degree of censorship imposed on media. As Figure 2.5 displayed, Hong Kong's press freedom has seen a vital, but generally negative development since the first ranking in 2002; starting on rank 18, three ranks above the UK and just one below the USA, it dropped to 56 in just one year and, after brief periods of recovery in 2004 and 2005 as well as in 2010, is currently at its all-time low on rank 70. While it was close to or even above the USA, the UK and Australia until 2005, the gap has since grown almost continuously and the region is now set apart clearly (Reporters without Borders 2015). In addition to the censorship imposed on media producers by external actors, Hong Kong news has been found to be prone to self-censorship (Lee 1998; Lee and Lin 2006; Lee and Chan 2008), which is a means of avoiding conflicts and external censorship from the start. Low degrees of evaluative content and frequent use of passive constructions and nominalisations, which avoid naming involved parties or shift responsibilities to other actors, reflect these restrictions on news production. Furthermore, Hong Kong newspapers contain the lowest range of variation in their language (see Figure 5.30), which shows that the degree of standardisation is highest here and underlines that the cause is not to be found in the style of one newspaper or author, but in more general aspects underlying the society.

While these results therefore confirm what was discussed in section 2.4, another assumption, namely the importance of international references particularly in economic news, has to be rejected. As was described above and mentioned already in section 6.1, international place names are relevant in news from Hong Kong, but in total numbers, Hong Kong stands out as having very small differences between the uses of international and national place names. The only domain in which international ones are significantly frequent is politics, a result which is mirrored in the UK and the USA. In economy on the other hand, Hong Kong shows a clear trend towards national matters, which is also compatible with the region's focus on economy, but suggests that here, too, proximity is a relevant factor and general news on economic matters around the world is considered less important than news which directly involves Hong Kong, even if these stories are on a smaller scale.

In contrast to Hong Kong, news from Kenya appears much less standardised in their language. This is surprising as especially here, English is an official, but not a dominant language; while in Hong Kong, proficiency in English has a traditional implication of professional success and was supported by the British during the time of colonialisation, it was kept intentionally low in Kenya and has therefore not been established as a necessity to such a degree, although the importance of English in the country is constantly increasing (Schneider 2007, 193). This would have suggested functional variation within Kenyan English to be limited and, again in combination with a low press freedom, for texts to be standardised heavily, which is not confirmed by the ranges the variety displays. Nevertheless, Kenyan news does reflect the issue of censorship in their overall values; like news from Hong Kong, it constructs a high distance and contains a lower level of evaluation than the native varieties, creating the notion of objectivity and avoiding personal statements and interpersonal aspects.

The two L2 varieties therefore differ in their degree of standardisation regarding the overall collection of investigated features, however, they both display an orientation towards written style. The results for the subdimension of medium indicate this most clearly for Hong Kong, which contained fewest pronouns, most nouns and the highest average lexical density. For the other two operationalisations, sentence-initial coordinators and nominalisations, Kenya tended most strongly towards the written medium, with Hong Kong following in second place. This attachment to the written medium for texts conveyed via the graphic channel reflects that English is not the dominant medium of everyday communication in these regions, but is more frequently found in restricted, official contexts in which the written channel and medium are prevailing (Schneider 2007, 196). Mesthrie and Bhatt (2008, 114), referring to studies on Indian and Malaysian English by Kachru (1983d) and Baskaran (1994), come to the conclusion that in many advanced and endonormatively stabilised varieties, written norms can be observed to be highly influential on language in general; although these two varieties differ in many aspects from the ones analysed here, this phenomenon is observable in the present study as well and thus appears to be more widely applicable.

Which native variety appeared closest to Kenya and Hong Kong in terms of medium varies slightly between the features, yet summarising the results indicates that the USA and Australia are most similar, whereas news from the UK produces the most spoken results for all linguistic markers. Taking into consideration the other parameters of tenor and field, Australia is positioned further apart from Kenya and Hong Kong. Especially regarding tenor, it clusters more clearly with the UK as both produce indicators for a low distance between author and reader. The USA on the other hand remains moderate throughout most features and can therefore be seen as the closest reference point for the L2 varieties in general. For the UK, the assumption formulated in section 2.4, namely a focus on entertainment and interpersonal aspects, can be said to have been confirmed; in addition to the tendency towards spoken style, the variety contains relatively many modals and evaluative adverbs and adjectives as well as direct addresses in the form of interrogatives and imperatives and 1st and 2nd person pronouns. Especially the interrogatives gave an insight into the particular phenomenon of puns in the British press which are frequently used independently of the seriousness or lightness of the topic, but can hardly be found in the other varieties.

News from the UK therefore appears very different from those of the L2 varieties. Articles from Australia and especially the USA, as was said above, are closer to Kenya and Hong Kong regarding distance and formality. Despite these similarities however, the varieties from Australia and the USA show a particularity in their use of place names. It was formulated in section 2.4 that due to the physical sizes of the countries, regional references were to be expected more frequently than in the other varieties. This assumption is reflected in the amount of research which has been conducted into local and regional news in the regions (e.g. Ramson 1989; Coutin and Chock 1996; Akhavan-Majid and Ramaprasad 1998; Kirkpatrick 2001; Bowd 2003; Duckworth et al. 2003; see section 2.3.3) and was confirmed in the present analysis. National names, especially those of prominent cities and states, showed to be very dominant not just in raw frequencies of their cumulative hits, but primarily in their numbers – among the twenty most frequent place names (see Table 5.15) Australia and the USA contained 11 and 10 national ones respectively, while the UK contained just seven and Kenya and Hong Kong five each. Although in both cases the name of the country is the most frequent item, a narrower scope on individual regions is therefore common and can be explained by an increased relevance of information for the respective society in contrast to irrelevance for the country as a whole; topics such as strikes, accidents and weather warnings are often limited in their impact to certain cities or regions, and also local politics is represented more strongly in Australia and the USA, indicating that a closer connection between audience and local matters is constructed in countries where the national level is so broad.

Regarding the L2 varieties in terms of their developmental status, the results in comparison to the native regions, even when interpreted independent of the individual domains and the functional variation within the varieties, suggest that the categorisation of Englishes in terms of a model is difficult and prone to obscuring differences. In the earlier taxonomies by McArthur (1987, 11) and Görlach (1990, 42), Kenyan and Hong Kong English can be found in the outer rings of the respective wheels, and Kachru (e.g. 1996, 71; 2006, 197) positions both varieties in his outer circle as norm-developing. Schneider (2007, 133-39, 189-97), too, categorises Kenyan and Hong Kong English together as both being in phase 3, although he predicts different developments for the two cases; for Kenya, he assumes that English is unlikely to become a symbol of national identity and that instead, a mixture of support and opposition from within the country will render Kenyan English endonormative, but rather homogeneous. In Hong Kong on the other hand, he points out that the future of the language depends on "economic, sociological, and political developments" of both Hong Kong and China, which makes it particularly hard to predict (ibid., 139). Overall, the status of English is secure due to its importance for business, which became apparent in parents' protests against the region's endeavours to decrease English-language education (Pennington and Yue 1994, 2; Bolton 2000, 271; Chan 2002, 271; see section 2.1.3). Especially the influence of China is argued to potentially trigger an anti-reaction in this context, should it try to impose a political direction which people in Hong Kong are unwilling to follow; in this case, English could become a distinguishing factor and a symbol of identity (Joseph 2004, 159-61; Schneider 2007, 139).

The results found for this study cannot confirm or falsify these general hypotheses regarding the future of the two varieties, as a diachronic perspective would be required to trace trends. Nevertheless, they demonstrate that there are differences between Kenyan and Hong Kong English news which are not in line with approaches which categorise them alike. Especially the aspect of homogeneity in Kenya has to be contradicted for the present study, as the variety demonstrates more internal variation as represented by the cumulative ranges than Hong Kong and even the UK. While the high values for Australia and the USA in this respect can again be explained by the size and spread of the population and the consequentially likely formation of different accents and dialects, Kenya most resembles the UK in population and size. The lower level of infrastructure and implicit restrictions in communication via telephone or digital channels might be a factor for a stronger differentiation of regional English varieties; however, Kenya also is home to dozens of native languages which are dominant in different areas and have varying influences on English, which certainly accounts for large parts of the country-internal variation and will continue to do so. This naturally restricts the level of homogeneity English can reach in Kenya, whereas in Hong Kong, the influence of native contact languages is limited to six (Ethnologue 2016), which might lead to a closer attachment to native varieties, or an increasingly characteristic, but less varying, Hong Kong English.

The variation in Kenya is mainly due to those domains which stood out across the analyses, namely lifestyle and hard news. The extreme values these domains produced for many linguistic features are largely responsible for the ranges of variation in all varieties, yet in Kenya, especially lifestyle expresses its characteristics very strongly when compared to the other news areas. This can be visualised best by looking at the deviation the domains display from the variety's average for the analysed linguistic features; Table 6.1 summarises these deviations in the cumulative values per domain (see appendix 4 for a detailed list of deviations for all features).

Table 6.1: Cumulative deviation from average values

	Australia	Hong Kong	Kenya	UK	USA
Economy	48.77	44.87	98.67	76.07	73.64
Hard News	176.92	174.46	161.84	106.89	98.12
Lifestyle	196.13	226.50	340.69	154.57	177.14
Politics	10.83	39.18	29.77	43.04	47.85
Sports	39.53	19.24	94.79	56.76	39.74

With regard to Kenya, the figures indicate that the domain of lifestyle is exempt from the strict and restrictive rules applying to news in general. When comparing lifestyle articles from Kenya to those of the other regions, we can find four names of politicians and the item *government* among the top twenty keywords, while the domain focuses more on health and leisure activities in the other varieties. Political and societal topics appear to be more frequently treated in this domain in Kenya, and the human interest angle might work as a less regulated outlet for opinions and critical statements that would otherwise be censored. In this context it is interesting to note that the degree of distinction of lifestyle news can be traced back to a large degree to the average length of the articles; in no other region is the difference between lifestyle and the rest so pronounced in this aspect as in Kenya, which underlines the low standardisation of the domain and indicates that critical topics are embedded in larger contexts, thereby broadening the focus and taking the edge off individual statements.

In terms of figures, a similar mechanism is visible in Hong Kong, where lifestyle is again the most distinct domain, yet the keyword list for this region does not indicate a strong coverage of political topics. This suggests that here, lifestyle news stories are more liberal and less strictly structured mainly because they contain less critical content. In the other domains, Hong Kong appears less specialised than Kenyan news, which display fairly high degrees of variation also with regard to sports and economic items. For the latter, this trend is mirrored in the USA and the UK, albeit to a lower degree. In Hong Kong on the other hand, this domain is the one closest to the average values; it displays characteristics mainly with regard to a low use of past tense, passive voice and nouns as well as a low lexical density, whereas in the other regions, it is outstanding also due to the articles notably differing in length.

In terms of the developmental status it therefore has to be concluded that Kenyan English is not only more diverse, but has developed characteristic registers for very specific and fine-tuned societal purposes. As lifestyle news in general has been found to contain a low degree of specialisation and aim at a broad target group, it can be assumed that in Kenya, this code, as well as its differences from that used in other news, is recognisable by many among the readers, and that the domain is furthermore accepted as a platform not just for soft news, but for political and sociocritical topics as well. This function is not just mirrored in the variety's isolated figures, but represented also in the distinction of lifestyle articles in Kenya in comparison to the native varieties. Since the Englishes from Australia, the USA and the UK are the dominant languages in their respective regions and the first choice for communication on all topics for most people living there, they can safely be assumed to have formed countless registers and display characteristics in their use depending on the context. Yet neither of the three shows as much deviation from the average values in any domain as Kenyan news does in lifestyle, which underlines the curious and particular position this domain holds here. In the native varieties an outsourcing of evaluative content is not necessary, which sets the domain apart regarding its subject matter, but makes the differences in terms of tenor of discourse more moderate, resulting in a lower internal variation.

The deviations from the average values referred to here of course do not hold any implications about the overall trends of the varieties, but only present another perspective on the internal variation. It has already been said that the L2 varieties generally display lower distances and more formal style, whereas the native varieties are more liberal in these aspects. This freedom is expressed in high internal ranges in Australia and the USA, however, in the latter the lower deviations show that the range is mainly caused by the domains of lifestyle, hard news and, to a certain degree, economy. In general, the distribution among US news is more centralised and the domains cluster more around the average value than in Australia, which displays more extreme deviations. Although the focus of the analysis was put on the L2 varieties and their developmental stages and characteristics, particularities for the native varieties which reflect the assumptions made earlier can thus be found and in many aspects and represent the different news landscapes and traditions as well as societal structures described in previous research.

7 Conclusion

On the basis of these results and the insights gained from them, this study answered as well as opened questions in the areas of variational linguistics and media studies, and brought to light some methodological advantages and concerns. The final sections will summarise and reflect on these dimensions and point out potential research fields and gaps for the future.

7.1 Summary of the Findings – "There and back again"

This study started out at a crossroad of three very different fields of linguistic research, variational linguistics, systemic functional linguistics and media language. Its aim was to conduct an analysis which would contribute to the understanding of all three, gaining insights into newspaper language, the status of regional varieties and the application of a functional approach to these different forms of variation.

To lay the foundation for the investigation, the study first discussed the individual frameworks and entangled different strands and usages of terminology, before describing the concrete situations and research traditions for each field and each of the five included regions. In this context, models depicting the linguistic variation of the English language were discussed and put into comparison, showing that although they represent varying degrees of detail, a categorisation of varieties, especially of so-called New Englishes, is difficult within the enclosed space of a model.

Chapter 3 then turned to the preparation of the analysis in terms of an operationalisation of the systemic functional concept of register. The discussion here lay open disputes about the exact parameters and subdimensions which are to reflect the context of situation, and summarised studies and operationalisations previously conducted into this area. While the suggestions for linguistic features suitable for an analysis were numerous, for this study many were redefined more concretely for the application to newspaper language. In terms of exact features, this implies that a one-to-one comparison to other research is restricted and the linguistic markers used here are not necessarily appropriate for another discourse field; yet universal operationalisations would be limited to very few features, and we argued that for a concise analysis of a particular discourse, specific functions of linguistic characteristics and means have to be taken into account in order to represent the parameters of register properly.

Having defined the exact linguistic markers to be analysed, chapter 4 introduced the dataset which formed the basis of the investigation. The corpus, which consists of 4,000 newspaper articles, was constructed along the principle of even stratification, meaning that all domains and varieties are represented equally instead of in their real-life proportions. The measurement of standardisation was the number of texts, not words, which rendered the parts diverse in overall tokens but allowed for all articles to be included as a whole, ensuring that results were not skewed due to a misrepresentation or lack of certain structural elements. The factual length of the individual texts as well the overall domains were incorporated in the analysis, where relative frequencies formed the bases for the discussions.

The results described and discussed in chapters 5 and 6 generally reflected two directions of comparison, namely among domains and among varieties, and from both perspectives, differences and similarities between the individual categories could be found. With regard to the news domains, the variation of which posed the first research question formulated in the introduction, a rough distinction between formal and less formal topics could be observed, a scale which was often delimited by hard news on the one and lifestyle on the other end. These two domains as well as sports news were most clearly set apart, while news from economy and politics appeared more strongly related to each other.

These general trends were found to be prominent in the analyses and the linguistic characteristics can therefore be concluded to be more domain- than region-dependent; however, the range of variation between the domains and the degree of specificity of individual ones differed greatly between the varieties and reflect the societies' media background and traditions. In this sense, the L2 varieties showed a clear tendency towards a more written style and a more strongly constructed distance between author and reader, confirming what has been found previously that written norms are a dominant influence in regions where English is primarily spoken as a second language (Mesthrie and Bhatt 2008, 114), and relating closely to their relatively low degrees of press freedom. Yet Kenya and Hong Kong did not align in all aspects, but showed differences especially in the differentiation of the domains and the internal ranges. In contrast to Hong Kong, which appeared most standardised and consistent with basic journalistic guidelines, Kenya displayed a specification particularly in lifestyle news, thus producing a comparatively high internal variation.

Such differences between the varieties in terms of functional variation are not only difficult to represent in a model, but are also very fine-grained and would most likely prove elusive in a study relying on a broader dataset like the ICE. A discourse-specific corpus provides more concise results and makes it possible to filter such differences, yet of course also implies limitations. When discussing the results at hand, it has to be kept in mind that they are particular to the field of news discourse and statements made about the developmental status are projections based on this very angle. Although the language used in this context reaches many people and is influential on language in general (Fowler 1991, 7; Bell 1998, 65; Conboy 2007, 4), it cannot claim to represent an entire variety. At this point, the approach to determine regional variation via functional variation either has to leave the detailed level of individual discourses, thereby losing exactly that valuable focus, or rely on a range of studies being conducted into different areas to create a pool of results for comparison. This not only requires more empirical work however, but leads back to the theories we started out from and calls for the clarification of some concepts so far left unconsidered, proving once more that the application of a theory is ultimately a way of pushing it forward (Wegener 2011, 215).

In the case of the assumption taken here that functional variation reflects the developmental status of a regional variety, a comparison of future research into other discourses would lead to the question of how broad or narrow the concept of "developmental status" is to be seen. For the present time, it was taken as an abstract value that can be determined for a variety and makes it comparable to others. In this sense it stayed on the level on which models are located, namely in considering the whole variety as one entity. However, it was shown in the analysis at hand that language is strongly affected by societal, yet discourse-specific factors like the press freedom, which makes it likely, if not unavoidable, that varieties will differ in the functional variation display depending on the discourse under investigation. While for newspaper language Kenyan English has to be regarded as more broadly developed than Hong Kong English, results for a different discourse field can lead to the opposite conclusion. The same holds true for the relation they display towards the native varieties, and even for comparisons among these.

Regarding the field of media language, the results described here trigger questions as well. Numerous studies have been conducted into the area of sports news, setting it apart quite clearly (e.g. Ghadessy 1988; Beard 1998; Dreyfus and Jones 2010; Aull and Brown 2013; contributions in Caldwell, Walsh and Vine, forthcoming). In studies dealing with more than one domain, soft and hard news are often distinguished, whereas the research questions are narrowed down to precise linguistic markers or ideological issues (e.g. Thompson, White and Kitley 2008; Reinemann et al. 2012; see section 2.3.3). Broad comparisons are rare, and a distinction between domains within the areas of soft and hard news has largely been neglected, yet looking at the linguistic particularities that were found in this study, the existence of conventions and norms can hardly be denied.

From a systemic functional perspective, this raises the question of whether or not these domains can be treated as separate registers. The answer to this depends on the exact definition, of which, as was seen in section 2.2, there are numerous. For the purpose of corpus compilation, the domains were defined language-externally in this study, relying on the categorisation given by the newspapers themselves. It was seen that the domains differed in several linguistic markers, and especially the parameter of field of discourse and its subdimension of experiential domain resulted in a clear untangling of the domains. Regarding the contextual configuration defined by Halliday and Hasan (e.g. Halliday 1974, 50; Halliday and Hasan 1985, 55-56), they therefore differ crucially and have to be counted as separate registers. Nevertheless, the predefined categories mean that individual texts were mainly considered within their domains and counted to the average values; individual exceptions were thus depicted as outliers, but an inductive approach such as Biber's (e.g. 1985, 1988, 1995) would offer an interesting additional perspective on the data to see whether on a linguistic basis, other clusters would emerge than suggested by the newspapers.

These trains of thought necessarily lead to some more detailed reflections on the method applied here. Like the circuit leading back to considerations about the theoretical frameworks, the analyses and results did not only give insights into the data, but also indicated advantages and disadvantages of the operationalisation and its usage. The following section will therefore summarise these aspects and make suggestions for future development of the method.

7.2 Reflections on the Method

In general, as register is defined as language in use (e.g. Halliday, McIntosh and Strevens 1964, 87; Halliday 1974, 32), it can be said that an empirical approach is most true to the concept itself. As the present study aimed at analysing both regional and functional variation on a larger scale, a quantitative approach was most suitable as it allows the comparison of the contextual configuration of large sets of texts as well as categories (Lukin et al. 2011, 206; Neumann 2013, 3).

Quantitative analyses of course always bring with them the need to generalise; for every linguistic feature investigated in this study, the results may contain instances in which the feature is used differently or in a context not relevant at the specific point. These faulty hits are eclipsed by the mass of results and could only be filtered out by an additional qualitative analysis, which is often impossible due to the size of the dataset. The design of the present study, in which linguistic features were not investigated at random but in relation to a concrete subdimension of one contextual parameter, counteracts this to a certain degree; of course, the results are still not perfectly accurate and contain irrelevant instances, yet since every feature was analysed with the focus on one specific function relating to one clearly defined aspect of the context, problematic features were more easily traceable. One such case in the analysis conducted here were interrogatives as part of the subdimension of social role relationship, which displayed various functions and usages in the dataset and were found to be unassignable to constructing either equal or unequal relations between reader and author; instead, which function they were intended to fulfil depended heavily on the exact context and would require an analysis of every individual instance.

This example indicates one problem of the concept of operationalisation, namely that it is difficult to account for different, sometimes even oppositional functions one linguistic feature can potentially take within one subdimension. Although not many such extreme cases as interrogatives were found, quantifying these would mean a very high error rate and a questionable interpretation. Instead, they are valuable as indicators of starting points for future research – a linguistic characteristic which is used in such diverse ways is certainly worth investigating in more detail in order to understand the discourse strategies at work and the implications they have.

Apart from having multiple functions within one subdimension, a linguistic feature can of course also represent different dimensions of context. This was observable for a number of features investigated here and is logical seeing that the three parameters influence each other and have many points of contact. As long as the current focus is kept in mind when discussing specific results, this multiple use does not have any disadvantages for the analysis itself, however, it does have implications for the statistical representation. Ideally, all linguistic features subsumed under one subdimension would show a correlation in the end, indicating a clear trend and proving their assignment as appropriate. In practice, this is rather unrealistic, not least due to the potentially diverse functions of a feature within one dimension discussed above. A multiple operationalisation of one feature weakens this possibility further, as a strong correlation in all subdimensions in which it is used would make the distinction between the subdimensions redundant.

Despite this complication, it is of course necessary to include a linguistic feature for all subdimensions for which it promises or has been found to be representative. Limiting every feature to just one parameter would mean ignoring the complexity of language and its users. The issue of multiple inclusion is of course not given in an inductive approach, for which features are not analysed in the context of concrete parameters but are pooled and used to let patterns emerge independent of predefined contextual aspects. Despite this advantage, as mentioned earlier, such an analysis does not offer the possibility to trace the differences between texts back to their source in the discourse situation; clear assignments to the parameters of field, tenor and mode on the other hand have this potential, and the intermediate step of their respective subdimensions ensures that the transition from abstract variables of context to concrete linguistic markers is comprehensible and reproducible.

Because of this theoretically guided nature, it has to be said that approaching data via an operationalisation rather than inductively has proven valuable for an analysis into regional variation. The developmental status of a variety does not only depend on general differences between text domains, but has to be considered more closely taking into account where the differences come from. From this perspective, trends become evident which relativise or emphasise regional or topical differences, such as a tendency towards informal and spoken style in lifestyle and sports news, or towards formal and written medium in the varieties from Kenya and Hong Kong, and interpretations of these differences between varieties and their societies can be more detailed and to the point. Thus, although a deductive approach implies limitations and cannot avoid the issues mentioned above, these disadvantages are outweighed by the concise nature of the results and make it an excellent tool for the application field of variational linguistics and the question of developmental status.

7.3 Future Directions and Open Questions

The present study is, to our knowledge, the first large-scale investigation into different newspaper domains across different regional varieties of English, and therefore not only provides results and answers, but also has limitations which offer themselves as anchor points for future research. From the perspective of media language, some very detailed aspects like the use of functionally diverse linguistic features would certainly be of interest, as would be a comparison between web and print news. As was stated in section 4.2, the present analysis drew exclusively on articles published on the newspapers' websites, meaning that a comparison to print editions was not possible. In general, studies have been conducted into this aspect (e.g. Lewis 2003; Flew 2009; Kornetzki 2012), yet the inclusion of news from different regions would add a dimension which has so far been largely neglected, albeit it has to be said that differences especially regarding multimodality cannot be said to be purely linguistic, but also depend on technological availability and infrastructure.

Another broader issue which is strongly connected to extra-linguistic aspects is suggested by the correlations of language and press freedom found in this study. The connection was very striking, yet since the corpus was not compiled for diachronic use, it does not contain different years of newspaper writing in a representative way. Furthermore, the press freedom ranking produced by Reporters without Borders was only started in 2002, which limits the data points available for such an investigation. Nevertheless, with an updated ranking published every year, a longitudinal study of its reflection in newspaper language would be very promising.

Focusing on the varieties and the question of the developmental status, it would of course be desirable to include more. In a first step, a larger number of L2 varieties would be a valuable addition to get a clearer idea of universal trends and particularities. Seeing the curious position of lifestyle news in Kenya, it would furthermore be interesting to add other newspaper domains, especially those intended to contain evaluation, like opinion sections and editorials, to see whether other varieties or domains display a similar function and isolation. In a very advanced step, EFL-varieties could serve as further reference points, thereby also addressing the issue of the linguistic differences between English as a foreign and a second language. Enhancing the scale is only one direction in which to head though; looking at individual varieties, a closer analysis of any of the three parameters as well as their subdimensions still holds a lot of potential for more detailed insights, particularly when complemented by more elaborate qualitative analyses.

Finally, as has already been addressed in section 7.1, a definition of the status of a variety requires not only more specific results in one discourse, but a comparison of functional variation across discourse fields. This is of course a very ambitious goal and is made even more so as language as well as its discourses are continuously subject to change; nevertheless, every result adds to the bigger picture of the diversity of English, and the broad basis of previous theoretical and empirical research has set the stage for numerous and efficient potential future explorations.

Works Cited

Abbott, Gerry. 1981. "Editorial." *World Englishes* 1 (1): 1–4. doi:10.1111/j.1467-971X.1981.tb00438.x.

Abdulaziz, Mohamed H. 1982. "Patterns of Language Acquisition and Use in Kenya: Rural-Urban Differences." *International Journal of the Sociology of Language* 34:95–120. doi:10.1515/ijsl.1982.34.95.

———. 1991. "East Africa (Tanzania and Kenya)." In *English around the World: Sociolinguistic Perspectives*, edited by Jenny Cheshire, 391–404. Cambridge: Cambridge University Press.

Abercrombie, David. 1951. *The Way People Speak*. The Listener, September 6, 1951. London: BBC. Reprinted as: Abercrombie, David. 1965. "R.P. and local Accent". In *Studies in Phonetics and Linguistics*. London: Oxford University Press.

Adewunmi, Bim. 2013. "The sexual politics of hair – are all cuts created equally?" *The Guardian*, January 22. Accessed November 08, 2016. http://www.guardian.co.uk/commentisfree/2013/jan/22/sexual-politics-hair-cuts-denmark.

Aguilar-Sánchez, Jorge. 2005. "English in Costa Rica." *World Englishes* 24 (2): 161–72. doi:10.1111/j.1467-971X.2005.00401.x.

Ajala, Adekunle. 1983. "The Nature of African Boundaries." *Africa Spectrum* 18 (2): 177–89.

Akhavan-Majid, Roya, and Jyotika Ramaprasad. 1998. "Framing and Ideology: A Comparative Analysis of U.S. and Chinese Newspaper Coverage of the Fourth United Nations Conference on Women on the NGO Forum." *Mass Communication and Society* 1 (3-4): 131–52. doi:10.1080/15205436.1998.9677853.

Alderson, Albert W. 1908. *The Extinction in Perpetuity of Armaments and War.* London: P. S. King & Son. Accessed May 03, 2015. https://ia902607.us.archive.org/0/items/extinctioninperp00aldeuoft/extinctioninperp00aldeuoft_bw.pdf.

Alliance for Audited Media. 2013. "Top 25 U.S. Newspapers for March 2013." Accessed November 18, 2015. http://auditedmedia.com/news/research-and-data/top-25-us-newspapers-for-march-2013.aspx.

Allsopp, Richard. 2003. *Dictionary of Caribbean English Usage.* Oxford: Oxford University Press.

Alméciga, Wilder Y. E., and Reid Evans. 2014. "Mentor Texts and the Coding of Academic Writing Structures: A Functional Approach." *HOW: A Colombian Journal for Teachers of English* 21 (2): 94–111.

Alonso-Almeida, Francisco. 2008. "The Middle English Medical Charm: Register, Genre and Text Type Variables." *Neuphilologische Mitteilungen* 109 (1): 9–38.

American National Corpus. 2015: American National Corpus Project. Accessed May 09, 2016. www.anc.org.

Amin, Abdinasir A., Dejan Zurovac, Beth B. Kangwana, Joanne Greenfield, Dorothy N. Otieno, Willis S. Akhwale, and Robert W. Snow. 2007. "The Challenges of Changing National Malaria Drug Policy to Artemisinin-Based Combinations in Kenya." *Malaria Journal* 6 (72). doi:10.1186/1475-2875-6-72.

Ansaldo, Umberto. 2009. "The Asian Typology of English: Theoretical and Methodological Considerations." *English World-Wide* 30 (2): 133–48.

Anthony, Lawrence. 2005. "AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom." In *International Professional Communication Conference (IPCC)*, 729–37.

Antilla, Liisa. 2005. “Climate of Scepticism: US Newspaper Coverage of the Science of Climate Change.” *Global Environmental Change* 15 (4): 338–52. doi:10.1016/j.gloenvcha.2005.08.003.

Archer, Dawn. 2006. “(Re)initiating Strategies: Judges and Defendants in Early Modern English Courtrooms.” *Journal of Historical Pragmatics* 7 (2): 181–211. doi:10.1075/jhp.7.2.03arc.

Asante, Mabel Y. 2012. “Variation in Subject-Verb Concord in Ghanaian English.” *World Englishes* 31 (2): 208–25. doi:10.1111/j.1467-971X.2012.01751.x.

Ashcroft, Bill, Gareth Griffiths, and Helen Tiffin. 1989. *The Empire Writes Back: Theory and Practice in Post-Colonial Literature.* London, New York: Routledge.

Atkins, Sue, Jeremy Clear, and Nicholas Ostler. 1992. “Corpus Design Criteria.” *Literary and Linguistic Computing* 7 (1): 1–16. doi:10.1093/llc/7.1.1.

Atkinson, Dwight. 1999. *Scientific Discourse in Sociohistorical Context: The Philosophical Transactions of the Royal Society of London, 1675-1975.* Mahwah, N.J., London: Lawrence Erlbaum.

Au, Jade S. K., Paul S. F. Yip, Cecilia L. W. Chan, and Frances Y. W. Law. 2004. “Newspaper Reporting of Suicide Cases in Hong Kong.” *Crisis* 25 (4): 161–68. doi:10.1027/0227-5910.25.4.161.

Audit Bureau of Circulations. 2013. “AdNews ABC Circulation Results Nov 2013.” Accessed February 04, 2016.
http://yaffacdn.s3.amazonaws.com/live/adnews/files/dmfile/ABCNEWS.pdf.

Aull, Laura L., and David W. Brown. 2013. “Fighting Words: A Corpus Analysis of Gender Representations in Sports Reportage.” *Corpora* 8 (1): 27–52. doi:10.3366/cor.2013.0033.

Austin, John L. (1962) 1973. *How to Do Things with Words.* 2nd ed. New York: Oxford University Press.

Australian Bureau of Statistics. 2015. “Australian Demographic Statistics, March Quarter 2015.” Accessed November 07, 2015. http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/DE9E6EA551822B41CA257EC900136154/$File/31010_mar%202015.pdf.

Australian Communications and Media Authority. 2012. “Television Sets in Australian Households 2011: Current Stock and Consumer Expectations About Replacing Television Sets.” Accessed November 08, 2015.
http://www.acma.gov.au/~/media/Research%20and%20Analysis/Research/pdf/Television%20sets%20in%20Australian%20households%202011.pdf.

Australian Government - Department of Foreign Affairs and Trade. 2015. “About Australia.” Accessed November 07, 2015. http://dfat.gov.au/about-australia/pages/about-australia.aspx.

Auwera, Johan van der, Dirk Noël, and Astrid de Wit. 2012. “The Diverging 'Need (To)'s of Asian Englishes.” In *Mapping Unity and Diversity World-Wide: Corpus-Based Studies of New Englishes*, edited by Marianne Hundt and Ulrike Gut, 55–76. Varieties of English around the World 43. Amsterdam: John Benjamins.

Avison, Shannon, and Michael Meadows. 2000. “Speaking and Hearing: Aboriginal Newspapers and the Public Sphere in Canada and Australia.” *Canadian Journal of Communication* 25 (3). http://www.cjc-online.ca/index.php/journal/article/view/1163/1082. Accessed December 19, 2015.

Axelsson, Margareta W. 1998. *Contraction in British Newspapers in the Late 20th Century.* Studia Anglistica Upsaliensia 102. Uppsala: Ubsaliensis S. Academiae.

Axler, Maria, Anson Yang, and Trudy Stevens. 1998. "Current Language Attitudes of Hong Kong Chinese Adolescents and Young Adults." In *Language in Hong Kong at Century's End*, edited by Martha Pennington, 329–38. Hong Kong: Hong Kong University Press.

Bailey, Richard W. 1985. "The Idea of World English." *English Today* 1 (1): 3–6. doi:10.1017/S0266078400013018.

———. 1992. *Images of English: A Cultural History of the Language.* Cambridge: Cambridge University Press.

Baker, Paul. 2006. *Using Corpora in Discourse Analysis.* London: Continuum.

———. 2010a. "Representations of Islam in British Broadsheet and Tabloid Newspapers 1999–2005." *Journal of Language and Politics* 9 (2): 310–38. doi:10.1075/jlp.9.2.07bak.

———. 2010b. *Sociolinguistics and Corpus Linguistics.* Edinburgh: Edinburgh University Press.

Baker, Paul, and Erez Levon. 2015. "Picking the Right Cherries? A Comparison of Corpus-Based and Qualitative Analyses of News Articles about Masculinity." *Discourse & Communication* 9 (2): 221–36. doi:10.1177/1750481314568542.

Baker, Paul, Costas Gabrielatos, Majid KhosraviNik, Michał Krzyzanowski, Tony McEnery, and Ruth Wodak. 2008. "A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press." *Discourse & Society* 19 (3): 273–306. doi:10.1177/0957926508088962.

Baker, Paul, Costas Gabrielatos, and Tony McEnery. 2013. "Sketching Muslims: A Corpus Driven Analysis of Representations Around the Word 'Muslim' in the British Press 1998-2009." *Applied Linguistics* 34 (3): 255–78. doi:10.1093/applin/ams048.

Balasubramanian, Chandrika. 2009a. "Circumstance Adverbials in Registers of Indian English." *World Englishes* 28 (4): 485–508. doi:10.1111/j.1467-971X.2009.01608.x.

———. 2009b. *Register Variation in Indian English.* Amsterdam: John Benjamins.

Bao, Zhiming, and Huaqing Hong. 2006. "Diglossia and Register Variation in Singapore English." *World Englishes* 25 (1): 105–14. doi:10.1111/j.0083-2919.2006.00449.x.

Barbieri, Federica. 2005. "Quotative Use in American English: A Corpus-Based, Cross-Register Comparison." *Journal of English Linguistics* 33 (3): 222–56. doi:10.1177/0075424205282667.

Barés, Guzman M., and Enric Llurda. 2013. "Internationalization of Business English Communication at University: A Threefold Needs Analysis." *Ibérica* 26: 151–69.

Bargiela-Chiappini, Francesca, and Sandra Harris. 1997. *Managing Language: The Discourse of Corporate Meetings.* Pragmatics & Beyond New Series 44. Amsterdam: John Benjamins.

Barnett, Emma. 2012. "For sleep's sake have a screen break before bed." *The Daily Telegraph*, August 30. Accessed November 10, 2016. http://www.telegraph.co.uk/technology/news/9508190/For-sleeps-sake-have-a-screen-break-before-bed.html.

Barnhurst, Kevin G. 2010. "News Geography & Monopoly: The Form of Reports on US Newspaper Internet Sites." *Journalism Studies* 3 (4): 477–89. doi:10.1080/14616700220000019218.

Barron, Anne. 2012. *Public Information Messages: A Contrastive Genre Analysis of State-Citizen Communication.* Pragmatics & Beyond New Series 222. Amsterdam: John Benjamins.

Bartsch, Sabine. 2009. "Corpus Studies of Register Variation: An Exploration of Academic Registers." *International Journal of English Studies* 20 (1): 105–24.

Baskaran, Loga. 1994. "The Malaysian English Mosaic." *English Today* 10 (1): 27. doi:10.1017/S0266078400000857.

Bateman, John. 2006. "Special Issue on Genre: Introduction." *Linguistics and the Human Sciences* 2 (2): 177–83. doi:10.1558/lhs.v2i2.177.

Bateman, John, Judy Delin, and Renate Henschel. 2007. "Mapping the Multimodal Genres of Traditional and Electronic Newspapers." In *New Directions in the Analysis of Multimodal Discourse*, edited by Terry D. Royce and Wendy L. Bowcher, 147–72. New Jersey: Lawrence Erlbaum.

Beam, Randal A. 1996. "How Perceived Environmental Uncertainty Influences the Marketing Orientation of U.S. Daily Newspapers." *Journalism & Mass Communication Quarterly* 73 (2): 285–303. doi:10.1177/107769909607300202.

Beard, Adrian. 1998. *The Language of Sport.* London: Routledge.

Beaugrande, Robert-Alain de, and Wolfgang Dressler. 1981. *Einführung in die Textlinguistik.* Konzepte der Sprach- und Literaturwissenschaft 28. Tübingen: Niemeyer.

Bednarek, Monika. 2006. *Evaluation in Media Discourse: Analysis of a Newspaper Corpus.* London: Continuum.

———. 2007. "Local Grammar and Register Variation: Explorations in Broadsheet and Tabloid Newspaper Discourse." *Empirical Language Research Journal* 1 (1). http://ejournals.org.uk/ELR/article/2007/1. Accessed January 06, 2014.

———. 2008. *Emotion Talk across Corpora.* Basingstoke: Palgrave Macmillan.

Bednarek, Monika, and Helen Caple. 2010. "Playing with Environmental Stories in the News - Good or Bad Practice?" *Discourse & Communication* 4 (1): 5–31. doi:10.1177/1750481309351206.

———. 2012a. "'Value Added': Language, Image and News Values." *Discourse, Context & Media* 1 (2-3): 103–13. doi:10.1016/j.dcm.2012.05.006.

———. 2012b. *News Discourse.* London: Continuum.

———. 2014. "Why Do News Values Matter? Towards a New Methodological Framework for Analysing News Discourse in Critical Discourse Analysis and Beyond." *Discourse & Society* 25 (2): 135–58. doi:10.1177/0957926513516041.

Begi, Nyakwara. 2014. "Use of Mother Tongue as a Language of Instruction in Early Years of School to Preserve the Kenyan Culture." *Journal of Education and Practice* 5 (3): 37–49.

Bell, Alexander M. 1888. *Science* 11 (260): 39–40. doi:10.1126/science.ns-11.260.39.

Bell, Allan. 1984a. "Good Copy - Bad News: The Syntax and Semantics of News Editing." In *Applied Sociolinguistics*, edited by Peter Trudgill, 73–116. Applied Language Studies. London: Academic Press.

———. 1984b. "Language Style as Audience Design." *Language in Society* 13 (2): 145–204. doi:10.1017/S004740450001037X.

———. 1991. *The Language of News Media.* Language in Society 17. Oxford: Blackwell.

———. 1998. "The Discourse of Structure of News Stories." In *Approaches to Media Discourse*, edited by Allan Bell and Peter Garrett, 64–104. Oxford: Blackwell.

Benamara, Farah, Carmine Cesarano, Antonio Picariello, Diego Reforgiato, and V. S. Subrahmanian. 2007. "Sentiment Analysis: Adjectives and Adverbs Are Better than Adjectives Alone." *Proceedings of the International Conference on Weblogs and Social Media*. http://www.icwsm.org/papers/3--Benamara-Cesarano-Picariello-Reforgiato-Subrahmanian.pdf. Accessed March 05, 2016.

Benjamin, Walter. 1991. “Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit.” In *Gesammelte Schriften, Band I-2*, edited by Walter Benjamin, Rolf Tiedemann, Hermann Schweppenhäuser, Hella Tiedemann-Bartels, Tillman Rexroth, and Charles Baudelaire, 435–69. Suhrkamp Taschenbuch Wissenschaft 931. Frankfurt am Main: Suhrkamp.

Bennett, Gillian, and Anne Rowbottom. 1998. “'Born a Lady, Died a Saint': The Deification of Diana in the Press and Popular Opinion in Britain.” *Fabula* 39 (3-4): 197–208. doi:10.1515/fabl.1998.39.3-4.197.

Bennett, Joe. 2007. “The Discursive Construction of British Identity in Six Daily Newspapers.” *Edinburgh Working Papers in Applied Linguistics* 15: 1–15.

Benson, Phil. 2002. “Hong Kong Words: Variation and Context.” In *Hong Kong English: Autonomy and Creativity*, edited by Kingsley Bolton, 161–70. Aberdeen: Hong Kong University Press.

Benson, Rodney, Mark Blach-Ørsten, Matthew Powers, Ida Willig, and Sandra V. Zambrano. 2012. “Media Systems Online and Off: Comparing the Form of News in the United States, Denmark, and France.” *Journal of Communication* 62 (1): 21–38. doi:10.1111/j.1460-2466.2011.01625.x.

Benveniste, Emile. 1971. *Problems in General Linguistics.* 2 vols. Coral Gables, Florida: University of Miami Press.

Bergs, Alexander. 2004. “Letters: A New Approach to Text Typology.” *Journal of Historical Pragmatics* 5 (2): 207–27. doi:10.1075/jhp.5.2.04ber.

Berns, Margie. 2005. “Expanding on the Expanding Circle: Where Do WE Go from Here?” *World Englishes* 24 (1): 85–93. doi:10.1111/j.0883-2919.2005.00389.x.

Bevitori, Cinzia. 2010. *Representations of Climate Change: News and Opinion Discourse in UK and US Quality Press: A Corpus-Assisted Discourse Study.* Studi interdisciplinari su traduzione, lingue e culture 11. Bologna: Bononia University Press.

Bhatia, Vijay K. 1993. *Analysing Genre: Language Use in Professional Settings.* London: Longman.

———. 1997. “Genre-Mixing in Academic Introductions.” *English for Specific Purposes* 16 (3): 181–95. doi:10.1016/S0889-4906(96)00039-7.

Biber, Douglas. 1985. “Investigating Macroscopic Textual Variation through Multifeature / Multidimensional Analyses.” *Linguistics* 23 (2): 337–60. doi:10.1515/ling.1985.23.2.337.

———. 1986. “Spoken and Written Textual Dimensions in English: Resolving the Contradictory Findings.” *Language* 62 (2): 384–414. doi:10.2307/414678.

———. 1988. *Variation across Speech and Writing.* Cambridge: Cambridge University Press.

———. 1989. “A Typology of English Texts.” *Linguistics* 27 (1): 3–44. doi:10.1515/ling.1989.27.1.3.

———. 1990. “Methodological Issues Regarding Corpus-Based Analyses of Linguistic Variation.” *Literary and Linguistic Computing* 5 (4): 257–69. doi:10.1093/llc/5.4.257.

———. 1993a. “Representativeness in Corpus Design.” *Literary and Linguistic Computing* 8 (4): 243–57. doi:10.1093/llc/8.4.243.

———. 1993b. “Using Register-Diversified Corpora for General Language Studies.” *Computational Linguistics* 19 (2): 219–41.

———. 1995. *Dimensions of Register Variation: A Cross-Linguistic Comparison.* Cambridge: Cambridge University Press.

———. 2006a. “Stance in Spoken and Written University Registers.” *Journal of English for Academic Purposes* 5 (2): 97–116. doi:10.1016/j.jeap.2006.05.001.

———. 2006b. *University Language: A Corpus-Based Study of Spoken and Written Registers.* Amsterdam: John Benjamins.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan M. Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English.* Harlow: Longman.

Biber, Douglas, Ulla Connor, and Thomas A. Upton. 2007. *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure.* Studies in Corpus Linguistics 28. Amsterdam, Philadelphia: John Benjamins.

Biber, Douglas, and Susan M. Conrad. 2009. *Register, Genre, and Style.* Cambridge, UK: Cambridge University Press.

Biber, Douglas, and Edward Finegan. 2001. "Diachronic Relations among Speech-Based and Written Registers in English." In *Variation in English: Multi-Dimensional Studies*, edited by Susan M. Conrad and Douglas Biber, 66–83. Harlow: Longman.

Biermeier, Thomas. 2008. *Word-Formation in New Englishes: a Corpus-Based Analysis.* Münster: LIT Verlag.

Biewer, Carolin. 2011. "Modal Auxiliaries in Second Language Varieties of English." In *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*, edited by Joybrato Mukherjee and Marianne Hundt, 7–33. Amsterdam: John Benjamins.

Black, Ian. 2001. "Russia hints at rethink on Nato." *The Guardian*, October 4. Accessed November 08, 2016. http://www.guardian.co.uk/world/2001/oct/04/afghanistan.russia1?INTCMP=SRCH.

Black, Rosemary. 2011. "People who wear glasses are seen as smarter and more professional, study finds." *Daily News*, January 3. Accessed November 10, 2016. http://www.nydailynews.com/life-style/people-wear-glasses-smarter-professional-study-finds-article-1.148074.

Bobda, Augustin Simo. 2000. "Research on New Enlishes: a Critical Review of Some Findings so Far with a Focus on Cameroon English." *Arbeiten aus Anglistik und Amerikanistik* 25 (1): 53–70.

———. 2003. "The Formation of Regional and National Features in African English Pronunciation: An Exploration of Some Non-Interference Factors." *English World-Wide* 24 (1): 17–42. doi:10.1075/eww.24.1.03sim.

———. 2009. "The Meaning of English Words across Cultures, with a Focus on Cameroon and Hong Kong." *Journal of Multilingual and Multicultural Development* 30 (5): 375–89. doi:10.1080/01434630802147882.

Bochenski, Natalie. 2013. "16-year-old charged over violent city robbery." *The Sydney Morning Herald*, January 20. Accessed November 10, 2016. http://www.smh.com.au/queensland/16yearold-charged-over-violent-city-robbery-20130120-2d101.html.

Bohner, Gerd. 2001. "Writing about Rape: Use of the Passive Voice and other Distancing Text Features as an Expression of Perceived Responsibility of the Victim." *British Journal of Social Psychology* 40 (4): 515–29. doi:10.1348/014466601164957.

Bolt, Philip, and Kingsley Bolton. 1996. "The International Corpus of English in Hong Kong." In *Comparing English Worldwide: The International Corpus of English*, edited by Sidney Greenbaum, 197–214. Oxford: Clarendon Press.

Bolton, Kingsley. 1999. "World Englishes - the Way We Were." *World Englishes* 18 (3): 393–414. doi:10.1111/1467-971X.t01-1-00152.

———. 2000. "The Sociolinguistics of Hong Kong and the Space for Hong Kong English." *World Englishes* 19 (3): 265–85. doi:10.1111/1467-971X.00179.

———. 2006. “World Englishes Today.” In *The Handbook of World Englishes*, edited by Braj B. Kachru, Yamuna Kachru, and Cecil L. Nelson, 240–69. Blackwell Handbooks in Linguistics. Malden, Mass: Blackwell.

Bolton, Kingsley, David Graddol, and Christiane Meierkord. 2011. “Towards Developmental World Englishes.” *World Englishes* 30 (4): 459–80. doi:10.1111/j.1467-971X.2011.01735.x.

Bolton, Kingsley, and Helen Kwok. 1990. “The Dynamics of the Hong Kong Accent: Social Identity and Sociolinguistic Description.” *Journal of Asian Pacific Communication* 1 (1): 147–72.

Bolton, Kingsley, and Shirley Lim. 2000. “Futures for Hong Kong English.” *World Englishes* 19 (3): 429–43. doi:10.1111/1467-971X.00191.

Bolton, Kingsley, Gerald Nelson, and Joseph Hung. 2002. “A Corpus-Based Study of Connectors in Student Writing: Research from the International Corpus of English in Hong Kong (ICE-HK).” *International Journal of Corpus Linguistics* 7 (2): 165–82. doi:10.1075/ijcl.7.2.02bol.

Bond, Michael H. 1985. “Language as a Carrier of Ethnic Stereotypes in Hong Kong.” *The Journal of Social Psychology* 125 (1): 53–62. doi:10.1080/00224545.1985.9713508.

Bonfadelli, Heinz. 2004. *Medienwirkungsforschung I: Grundlagen und theoretische Perspektiven.* 3rd ed. 2 vols. Konstanz: UVK Verl.-Ges.

Bongartz, Christiane M., and Sarah Buschfeld. 2011. “English in Cyprus: Second Language Variety or Learner English?” In *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*, edited by Joybrato Mukherjee and Marianne Hundt, 35–54. Amsterdam: John Benjamins.

Böttger, Claudia. 2007. “Lost in Translation? An Analysis of the Role of English as the 'Lingua Franca' of Multilingual Business Communication.” Verlag Dr. Kovač. htp://www.verlagdr-kovac.de/978-3-8300-2930-4.htm.

Bowcher, Wendy L. 1999. “Investigating Institutionalization in Context.” In *Text and Context in Functional Linguistics*, edited by Mohsen Ghadessy, 141–76. Current Issues in Linguistic Theory 169. Amsterdam: John Benjamins.

Bowd, Kathryn. 2003. “How Different Is 'different'? Australian Country Newspapers and Development Journalism.” *Asia Pacific Media Educator* 1 (14): 117–30.

Boyd, Andrew, Peter Stewart, and Ray Alexander. 2008. *Broadcast Journalism: Techniques of Radio & Television News.* 6th ed. Amsterdam: Elsevier.

Boykoff, Maxwell T., and Jules M. Boykoff. 2004. “Balance as Bias: Global Warming and the US Prestige Press.” *Global Environmental Change* 14 (2): 125–36. doi:10.1016/j.gloenvcha.2003.10.001.

———. 2007. “Climate Change and Journalistic Norms: A Case-Study of US Mass-Media Coverage.” *Geoforum* 38 (6): 1190–1204. doi:10.1016/j.geoforum.2007.01.008.

Bradley, Matt. 2013. “New Premier Is Named in Egypt.” *The Wall Street Journal Asia*, July 9.

Braendle, Frederick. 1910. *World-English, a New World Language: Veltlang, with English Words and English Grammar, Subject to the Limitations of the Phonetic Writing of Veltlang, Together with a Simple Phonetic World-Alphabet Seuastikon, Modeled After the Most Primeval Characters of Writing Ever Used by the Human Race.* Washington: Gibson Brothers. Accessed February 06, 2016.
http://babel.hathitrust.org/cgi/pt?id=mdp.39015005654754;view=1up;seq=7.

Breeze, Ruth. 2014. “Perspectives on North and South: The 2012 Financial Crisis in Spain Seen Through Two Major British Newspapers.” *Discourse & Communication* 8 (3): 241–59. doi:10.1177/1750481314537572.

Brinker, Klaus. (1985) 2005. *Linguistische Textanalyse: Eine Einführung in Grundbegriffe und Methoden.* 6th ed. Berlin: Schmidt.

Brook, George L. 1979. *Varieties of English.* 2nd ed. London: Palgrave Macmillan.

Brookes, Heather J. 1995. “'Suit, Tie and a Touch of Juju' - the Ideological Construction of Africa: A Critical Discourse Analysis of News on Africa in the British Press.” *Discourse & Society* 6 (4): 461–94. doi:10.1177/0957926595006004002.

Brown, Roger, and Albert Gilman. 1960. “The Pronouns of Power and Solidarity.” In *Style in Language*, edited by Thomas A. Seboek, 253–76. Cambridge, MA: MIT Press.

Brownlees, Nicholas. 2006. *News Discourse in Early Modern Britain: Selected Papers of CHINED 2004.* Linguistic Insights: Studies in Language and Communication 30. Bern: Peter Lang.

Bruthiaux, Paul. 1994. “Me Tarzan, You Jane: Linguistic Simplification in 'Personal Ads' Register.” In *Sociolinguistic Perspectives on Register*, edited by Douglas Biber and Edward Finegan. Oxford: Oxford University Press.

———. 1996. *The Discourse of Classified Advertising: Exploring the Nature of Linguistic Simplicity.* Oxford: Oxford University Press.

———. 2000. “In a Nutshell: Persuasion in the Spatially Constrained Language of Advertising.” *Language & Communication* 20 (4): 297–310. doi:10.1016/S0271-5309(00)00003-3.

Bryce, James. 1888. *The American Commonwealth* 3. London: Macmillan.

Brysbaert, Marc, Boris New, and Emmanuel Keuleers. 2012. “Adding Part-of-Speech Information to the SUBTLEX-US Word Frequencies.” *Behavior Research Methods* 44 (4): 991–97. doi:10.3758/s13428-012-0190-4.

Bubela, Tania M., and Timothy A. Caulfield. 2004. “Do the Print Media "Hype" Genetic Research? A Comparison of Newspaper Stories and Peer-Reviewed Research Papers.” *Canadian Medical Association Journal* 170 (9): 1399–1407. doi:10.1503/cmaj.1030762.

Bühler, Karl. (1934) 2011. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Stuttgart: Gustav Fischer Verlag. Translated by Donald Fraser Goodwin as *Theory of Language: The Representational Function of Language* (Amsterdam: John Benjamins, 2011).

Bull, Peter, and Anita Fetzer. 2006. “Who Are We and Who Are You? The Strategic Use of Forms of Address in Political Interviews.” *Text & Talk* 26 (1): 3–37. doi:10.1515/TEXT.2006.002.

Bunton, David. 2005. “The Structure of PhD Conclusion Chapters.” *Journal of English for Academic Purposes* 4 (3): 207–24. doi:10.1016/j.jeap.2005.03.004.

Buregeya, Alfred. 2006. “Grammatical Features of Kenyan English and the Extent of Their Acceptability.” *English World-Wide* 27 (2): 199–216.

Burger, Harald. 2005. *Mediensprache: Eine Einführung in Sprache und Kommunikationsformen der Massenmedien.* 3rd ed. Berlin: De Gruyter.

Butt, David, and Rebekah Wegener. 2007. “The Work of Concepts: Context and Metafunction in the Systemic Functional Model.” In *Continuing Discourse on Language: a Functional Perspective (Vol. 2)*, edited by Ruqaiya Hasan, Christian M. I. M. Matthiessen, and Jonathan Webster, 589–618. London: Equinox.

Caldwell, David, John Walsh, Elaina W. Vine, and Jon Jureidini, eds. Forthcoming. *The Discourse of Sport: Analyses from Social Linguistics.* Routledge Studies in Sociolinguistics. London, New York: Routledge.

Camus, Julia T. W. 2009. “Metaphors of Cancer in Scientific Popularization Articles in the British Press.” *Discourse Studies* 11 (4): 465–95. doi:10.1177/1461445609105220.

Canter, Lily. 2013. “The Misconception of Online Comment Threads: Content and Control on Local Newspaper Websites.” *Journalism Practice* 7 (5): 604–19. doi:10.1080/17512786.2012.740172.

Caple, Helen. 2010. “Doubling-Up: Allusion and Bonding in Multi-Semiotic News Stories.” In *New Discourse on Language: Functional Perspectives on Multimodality, Identity, and Affiliation*, edited by Monika Bednarek and James R. Martin, 111–33. London, New York: Continuum.

Caple, Helen, and Monika Bednarek. 2015. “Rethinking News Values: What a Discursive Approach Can Tell Us about the Construction of News Discourse and News Photography.” *Journalism* 2:1–22. doi:10.1177/1464884914568078.

Carless, David R. 1995. “Politicised Expressions in the South China Morning Post.” *English Today* 11 (2): 18. doi:10.1017/S026607840000821X.

Carpenter, Serena. 2007. “U.S. Elite and Non-Elite Newspapers' Portrayal of the Iraq War: A Comparison of Frames and Source Use.” *Journalism & Mass Communication Quarterly* 84 (4): 761–76. doi:10.1177/107769900708400407.

Carter, Felice. 1968. “The Press in Kenya.” *International Communication Gazette* 14 (2): 85–88. doi:10.1177/001654926801400203.

Carvalho, Anabela. 2005. “Representing the Politics of the Greenhouse Effect: Discursive Strategies in the British Media.” *Critical Discourse Studies* 2 (1): 1–29. doi:10.1080/17405900500052143.

Chafe, Wallace. 1992. “Discourse.” In *International Encyclopedia of Linguistics*, edited by William Bright, 355–58. Oxford: Oxford University Press.

Chambers, Jack K. 2004. “Dynamic Typology and Vernacular Universals.” In *Dialectology Meets Typology. Dialect Grammar from a Crosslinguistic Perspective*, edited by Bernd Kortmann, 127–45. Berlin: De Gruyter Mouton.

Chan, Brian H.-S. 2009. “English in Hong Kong Cantopop: Language Choice, Code-Switching and Genre.” *World Englishes* 28 (1): 107–29. doi:10.1111/j.1467-971X.2008.01572.x.

Chan, Elaine. 2002. “Beyond Pedagogy: Language and Identity in Post-Colonial Hong Kong.” *British Journal of Sociology of Education* 23 (2): 271–85.

Chan, Jim. 2013. “Contextual Variation and Hong Kong English.” *World Englishes* 32 (1): 54–74. doi:10.1111/weng.12004.

Chan, Joseph M., and Francis L. Lee. 2011. “The Primacy of Local Interests and Press Freedom in Hong Kong: A Survey Study of Professional Journalists.” *Journalism* 12 (1): 89–105. doi:10.1177/1464884910385189.

Chan, Michael. 2014. “(Re)categorizing Intergroup Relations and Social Identities Through News Discourse: The Case of the China Daily's Reporting on Regional Conflict.” *Journal of Language and Social Psychology* 33 (2): 144–64. doi:10.1177/0261927X13508125.

Chapman, Simon, and Amanda Dominello. 2001. “A Strategy for Increasing News Media Coverage of Tobacco and Health in Australia.” *Health Promotion International* 16 (2): 137–43. doi:10.1093/heapro/16.2.137.

Charles, Maggie. 2003. “‘This Mystery…’: A Corpus-Based Study of the Use of Nouns to Construct Stance in Theses from Two Contrasting Disciplines.” *Journal of English for Academic Purposes* 2 (4): 313–26. doi:10.1016/S1475-1585(03)00048-1.

———. 2006a. “Phraseological Patterns in Reporting Clauses Used in Citation: A Corpus-Based Study of Theses in Two Disciplines.” *English for Specific Purposes* 25 (3): 310–31. doi:10.1016/j.esp.2005.05.003.

———. 2006b. “The Construction of Stance in Reporting Clauses: A Cross-Disciplinary Study of Theses.” *Applied Linguistics* 27 (3): 492–518. doi:10.1093/applin/aml021.

Cheeseman, Nic. 2014. “Are Newspapers on Their Way out?” *Daily Nation*, April 30. Accessed January 26, 2016. http://www.nation.co.ke/oped/Opinion/Media-Newspapers-Kenya-Daily-Nation/-/440808/2262878/-/7udebv/-/index.html.

Chen, Lily. 2005. “Transitivity in Media Texts: Negative Verbal Process Sub-Functions and Narrator Bias.” *International Review of Applied Linguistics in Language Teaching* 43 (1): 33–51. doi:10.1515/iral.2005.43.1.33.

———. 2007. “Negatives and Positives in the Language of Politics: Attitudes towards Authority in the British and Chinese Press.” *Journal of Language and Politics* 6 (3): 475–501. doi:10.1075/jlp.6.3.12che.

Cheng, Winnie, and Phoenix Lam. 2010. “Media Discourses in Hong Kong: Change in Representation of Human Rights.” *Text & Talk* 30 (5): 507–27. doi:10.1515/text.2010.025.

Cheung, Anne S. Y. 2003. “Hong Kong Press Coverage of China - Taiwan Cross Straits Tension.” In *Hong Kong in Transition: One Country, Two Systems*, edited by Robert Ash, Peter Ferdinand, Brian Hook, and Robin Porter, 210–25. Studies in the Modern History of Asia 11. London, New York: Routledge.

Cheung, Yat-shing. 1985. “Power, Solidarity, and Luxury in Hong Kong: a Sociolinguistic Study.” *Anthropological Linguistics* 27 (2): 190–203.

Chevillet, François. 1993. “English or Englishes?” *English Today* 9 (4): 29–33. doi:10.1017/S0266078400007288.

Childs, Becky, and Walt Wolfram. 2004. “Bahamian English: Phonology.” In *A Handbook of Varieties of English*. Vol. 1, edited by Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie, and Clive Upton. 2 vols, 435–49. Berlin: De Gruyter Mouton.

China Daily. 2014. “China Daily Media Profile 2014.” Accessed November 12, 2015. http://www.chinadailyasia.com/about/our-products/HKedition-MediaKit.pdf.

Christie, Pauline. 1989. “Questions of Standards and Intra-Regional Differences in Caribbean Examinations.” In *English across Cultures, Cultures across English*, edited by Ofelia García and Ricardo Otheguy. Berlin: De Gruyter Mouton.

Christophersen, Paul. 1988. “'Native Speakers' and World English.” *English Today* 4 (3): 15–18. doi:10.1017/S0266078400003473.

Churchill, Winston. 1956-1958. *A History of the English-Speaking Peoples.* London: Cassell.

CIA (Central Intelligence Agency). 2016. “The World Factbook.” Accessed January 15, 2016. https://www.cia.gov/library/publications/resources/the-world-factbook/.

Claridge, Claudia. 2005. “Questions in Early Modern English Pamphlets.” *Journal of Historical Pragmatics* 6 (1): 133–68. doi:10.1075/jhp.6.1.07cla.

Clark, Caroline. 2010. “Evidence of Evidentiality in the Quality Press 1993 and 2005.” *Corpora* 5 (2): 139–60. doi:10.3366/cor.2010.0103.

Clegg-Smith, Katherine, Yvonne Terry-McElrath, Melanie Wakefield, and Russil Durrant. 2005. “Media Advocacy and Newspaper Coverage of Tobacco Issues: a Comparative Analysis of 1 Year's Print News in the United States and Australia.” *Nicotine & Tobacco Research* 7 (2): 289–99. doi:10.1080/14622200500056291.

Clegg-Smith, Katherine, Melanie Wakefield, and Elizabeth Edsall. 2006. “The Good News about Smoking: How Do U.S. Newspapers Cover Tobacco Issues?” *Journal of Public Health Policy* 27 (2): 166–81. doi:10.1057/palgrave.jphp.3200079.

Coetzee-Van Rooy, Susan, and Bertus van Rooy. 2005. “South African English: Labels, Comprehensibility and Status.” *World Englishes* 24 (1): 1–19. doi:10.1111/j.0883-2919.2005.00384.x.

Collins, Patricia A., Julia Abelson, Heather Pyman, and John N. Lavis. 2006. “Are We Expecting Too Much from Print Media? An Analysis of Newspaper Coverage of the 2002 Canadian Healthcare Reform Debate.” *Social Science & Medicine* 63 (1): 89–102. doi:10.1016/j.socscimed.2005.12.012.

Conboy, Martin. 2006. *Tabloid Britain: Constructing a Community through Language.* London: Routledge.

———. 2007. *The Language of the News.* London: Routledge.

Condit, Celeste M., and J. A. Selzer. 1985. “The Rhetoric of Objectivity in the Newspaper Coverage of a Murder Trial.” *Critical Studies in Mass Communication* 2 (3): 197–216. doi:10.1080/15295038509360081.

Conduit, Anne M. 1991. “Linguistic Choice and Public Opinion: the Australian Media and Indonesia/Australia Relation.” *Lingua* 2: 25–71.

Connor, Ulla, and Anna Mauranen. 1999. “Linguistic Analysis of Grant Proposals: European Union Research Grants.” *English for Specific Purposes* 18 (1): 47–62. doi:10.1016/S0889-4906(97)00026-4.

Conrad, Peter. 1999. “Uses of Expertise: Sources, Quotes, and Voice in the Reporting of Genetics in the News.” *Public Understanding of Science* 8 (4): 285–302. doi:10.1088/0963-6625/8/4/302.

Conrad, Susan M. 1996. “Investigating Academic Texts with Corpus-Based Techniques: An Example from Biology.” *Linguistics and Education* 8 (3): 299–326. doi:10.1016/S0898-5898(96)90025-X.

Conrad, Susan M., and Douglas Biber, eds. 2001. *Variation in English: Multi-Dimensional Studies.* Harlow: Longman.

Cooper, Rob. 2011. “Now 20 business leaders declare Cameron's veto a 'diplomatic coup' as Clegg vows to get own way at next summit.” *Daily Mail*, December 23. Accessed November 07, 2016. http://www.dailymail.co.uk/news/article-2077904/Nick-Clegg-tells-Lib-Dems-set-agenda-EU-summit.html.

Cormack, Mike. 1995. “The Use of Gaelic in Scottish Newspapers.” *Journal of Multilingual and Multicultural Development* 16 (4): 269–80. doi:10.1080/01434632.1995.9994607.

Coseriu, Eugenio. 1994. *Textlinguistik.* 3rd ed. Tübingen, Basel: Francke Verlag.

Cotter, Colleen. 2011. “Diversity Awareness and the Role of Language in Cultural Representations in News Stories.” *Journal of Pragmatics* 43 (7): 1890–99. doi:10.1016/j.pragma.2010.09.020.

Coupland, Nikolas, Hywel Bishop, and Peter Garrett. 2003. “Home Truths: Globalisation and the Iconising of Welsh in a Welsh-American Newspaper.” *Journal of Multilingual and Multicultural Development* 24 (3): 153–77. doi:10.1080/01434630308666495.

Coutin, Susan B., and Phyllis P. Chock. 1996. ““Your Friend, the Illegal": Definition and Paradox in Newspaper Accounts of U.S. Immigration Reform.” *Identities* 2 (1-2): 123–48. doi:10.1080/1070289X.1997.9962529.

Couture, Barbara. 1986. “Effective Ideation in Written Text: A Functional Approach to Clarity and Exigence.” In *Functional Approaches to Writing: Research Perspectives*, edited by Barbara Couture, 69–91. London: Pinter Publishers.

Craig, Dennis. 1983. “Toward a Description of Caribbean English.” In *The Other Tongue: English across Cultures*, edited by Braj B. Kachru. 1st ed., 198–209. Oxford: Pergamon Press.

Crawford, Ewan. 2009. “A New Sort of Democracy? The Opinion Pages in the Scottish Daily Quality Press.” *Journalism* 10 (4): 451–72. doi:10.1177/1464884909104951.

Crystal, David. 1985. “How Many Millions? The Statistics of English Today.” *English Today* 1 (1): 7–9. doi:10.1017/S026607840001302X.

———. 1987. *The Cambridge Encyclopedia of Language.* Cambridge: Cambridge University Press.

———. 1999. “The Future of Englishes.” *English Today* 15 (2): 10–20. doi:10.1017/S0266078400010816.

———. 2001. *Language and the Internet.* Cambridge: Cambridge University Press.

———. 2003. *English as a Global Language.* 2nd ed. Cambridge: Cambridge University Press.

Csomay, Eniko. 2002. “Variation in Academic Lectures: Interactivity and Level of Instruction.” In *Using Corpora to Explore Linguistic Variation*, edited by Randi Reppen, Susan M. Fitzmaurice, and Douglas Biber, 203–24. Studies in Corpus Linguistics 9. Amsterdam: John Benjamins.

Culpeper, Jonathan, and Merja Kytö. 2010. *Early Modern English Dialogues: Spoken Interaction as Writing.* Studies in English language. Cambridge: Cambridge University Press.

Cummings, Patrick J., and Hans-Georg Wolf. 2011. *A Dictionary of Hong Kong English: Words from the Fragrant Harbor.* Hong Kong: Hong Kong University Press.

Curran, James, Shanto Iyengar, Anker Brink Lund, and Inka Salovaara-Moring. 2009. “Media System, Public Knowledge and Democracy: A Comparative Study.” *European Journal of Communication* 24 (1): 5–26. doi:10.1177/0267323108098943.

Daily Nation. 2016. “About Daily Nation.” Accessed January 26, 2016. http://www.nation.co.ke/meta/-/1194/1172/-/ojmv8c/-/index.html.

Daneš, František. 1970. “Zur linguistischen Analyse der Textstruktur.” *Folia Linguistica* 4 (1-2): 72–78. doi:10.1515/flin.1970.4.1-2.72.

———. 1976. “Zur semantischen und thematischen Struktur des Kommunikats.” In *Probleme der Textgrammatik*, edited by František Daneš and Dieter Viehweger, 29–41. Berlin: Akademie-Verlag.

Danielson, Wayne, Dominic Lasorsa, and Dae Im. 1992. “Journalists and Novelists: a Study of Diverging Styles.” *Journalism Quarterly* 69 (2): 436–46. doi:10.1177/107769909206900217.

Davies, Mark. 2008. “The Corpus of Contemporary American English: 520 Million Words, 1990-present.” Accessed February 28, 2016. http://corpus.byu.edu/coca/.

del-Teso-Craviotto, Marisol. 2006. “Language and Sexuality in Spanish and English Dating Chats.” *Journal of Sociolinguistics* 10 (4): 460–80. doi:10.1111/j.1467-9841.2006.00288.x.

Deterding, David, Jennie Wong, and Andy Kirkpatrick. 2008. “The Pronunciation of Hong Kong English.” *English World-Wide* 29 (2): 148–75.

Devonish, Hubert, and Harry G. Otelemate. 2004. “Jamaican Creole and Jamaican English: Phonology.” In *A Handbook of Varieties of English*. Vol. 1, edited by Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie, and Clive Upton. 2 vols, 450–80. Berlin: De Gruyter Mouton.

Dijk, Teun van. 1980. *Macrostructures: An Interdisciplinary Study of Global Structures in Discourse, Interaction, and Cognition.* Hillsdale, N.J. Lawrence Erlbaum.

———. 1983. “Discourse Analysis: Its Development and Application to the Structure of News.” *Journal of Communication* 33 (2): 20–43. doi:10.1111/j.1460-2466.1983.tb02386.x.

———. 1984. *Structures of International News: A Case Study of the World's Press.* Report for UNESCO. Amsterdam: University of Amsterdam.

———. 1985. “Structures of News in the Press.” In *Discourse and Communication*, edited by Teun van Dijk, 69–93. Berlin: De Gruyter.

———. 1997. “The Study of Discourse.” In *Discourse as Structure and Process*, edited by Teun van Dijk, 1–34. London: Sage Publications.

———. 1998. “Opinions and Ideologies in the Press.” In *Approaches to Media Discourse*, edited by Allan Bell and Peter Garrett, 21–63. Oxford: Blackwell.

Dimitrova, Daniela V., and Jesper Strömbäck. 2005. “Mission Accomplished? Framing of the Iraq War in the Elite Newspapers in Sweden and the United States.” *International Communication Gazette* 67 (5): 399–417. doi:10.1177/0016549205056050.

Dirks, Una. 2006. “Critical Discourse Analysis of the Iraq Conflict in the British and German 'Quality' Press.” *Revista Alicantina de Estudios Ingleses* 19: 101–23.

Diwersy, Sascha, Stefan Evert, and Stella Neumann. 2014. “A Weakly Supervised Multivariate Approach to the Study of Language Variation.” In *Aggregating Dialectology, Typology, and Register Analysis: Linguistic Variation in Text and Speech*, edited by Benedikt Szmrecsanyi and Bernhard Wälchli. Lingua & Litterae 28. Berlin: De Gruyter.

Djankov, Simeon, Caralee McLiesh, Tatiana Nenova, and Andrei Shleifer. 2011. “Who Owns the Media?” Accessed November 20, 2015. http://www.nber.org/papers/w8288.pdf.

Dowding, Keith, and Chris Lewis. 2012. “Newspaper Reporting and Changing Perceptions of Ministerial Accountability in Australia.” *Australian Journal of Politics & History* 58 (2): 236–50. doi:10.1111/j.1467-8497.2012.01634.x.

Dowell, Ben. 2006. “Morning Star Faces Olympic Challenge.” *The Guardian*, August 3. Accessed November 14, 2015. http://www.theguardian.com/media/2006/aug/03/olympicsandthemedia.pressandpublishing.

Drasovean, Anda, and Caroline Tagg. 2015. “Evaluative Language and Its Solidarity-Building Role on TED.com: An Appraisal and Corpus Analysis.” *Language@Internet* 12. http://www.languageatinternet.org/articles/2015/drasovean.

Dressen, Dacia. 2003. “Geologists' Implicit Persuasive Strategies and the Construction of Evaluative Evidence.” *Journal of English for Academic Purposes* 2 (4): 273–90. doi:10.1016/S1475-1585(03)00046-8.

Dreyfus, Shoshana, and Sandra C. Jones. 2010. “Constructing Sports Stars: Appliable Linguistics and the Language of the Media.” In *Appliable Linguistics: Reclaiming the Place of Language in Linguistics*, edited by Ahmar Mahboob and Naomi Knight, 114–29. London: Continuum.

Du-Babcock, Bertha. 2013. “English as Business Lingua Franca: A Comparative Analysis of Communication Behavior and Strategies in Asian and European Contexts.” *Ibérica* 26: 99–130.

Duckworth, Kenneth, John H. Halpern, Russell K. Schutt, and Christopher Gillespie. 2003. “Use of Schizophrenia as a Metaphor in U.S. Newspapers.” *Psychiatric Services* 54 (10): 1402–4. doi:10.1176/appi.ps.54.10.1402.

Dudo, Anthony, Sharon Dunwoody, and Dietram Scheufele. 2011. “The Emergence of Nano News: Tracking Thematic Trends and Changes in U.S. Newspaper Coverage of Nanotechnology.” *Journalism & Mass Communication Quarterly* 88 (1): 55–75. doi:10.1177/107769901108800104.

Duguid, Alison. 2010a. “Investigating Anti and Some Reflections on Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS).” *Corpora* 5 (2): 191–220. doi:10.3366/cor.2010.0105.

———. 2010b. “Newspaper Discourse Informalisation: A Diachronic Comparison from Keywords.” *Corpora* 5 (2): 109–38. doi:10.3366/cor.2010.0102.

Dunning, Eric. 1999. *Sport Matters: Sociological Studies of Sport, Violence, and Civilization.* London: Routledge.

Durrant, Russil, Melanie Wakefield, Kim McLeod, Katherine Clegg-Smith, and Simon Chapman. 2003. “Tobacco in the News: An Analysis of Newspaper Coverage of Tobacco Issues in Australia, 2001.” *Tobacco Control* 12 (Suppl. II): ii75-ii81. doi:10.1136/tc.12.suppl_2.ii75.

Eastman, Susan T., and Andrew C. Billings. 2000. “Sportscasting and Sports Reporting: The Powder of Gender Bias.” *Journal of Sport & Social Issues*, no. 24: 192–213. doi:10.1177/0193723500242006.

Eco, Umberto. 1995. *The Search for the Perfect Language.* Oxford: Blackwell.

Egbert, Jesse. 2012. “Style in Nineteenth Century Fiction: A Multi-Dimensional Analysis.” *Scientific Study of Literature* 2 (2): 167–98. doi:10.1075/ssol.2.2.01egb.

Eggins, Suzanne. 1994. *An Introduction to Systemic Functional Linguistics.* New York: Continuum.

Eggins, Suzanne, and James R. Martin. 1997. “Genres and Registers of Discourse.” In *Discourse as Structure and Process*, edited by Teun van Dijk, 230–56. London: Sage Publications.

Ehlich, Konrad. 1980. “Der Alltag des Erzählens.” In *Erzählen im Alltag*, edited by Konrad Ehlich, 11–27. Frankfurt am Main: Suhrkamp.

———. 1986. “Funktional-pragmatische Kommunikationsanalyse – Ziele und Verfahren.” In *Untersuchungen zur Kommunikation: Ergebnisse und Perspektiven*, edited by Wolf-Dietrich Hartung, 15–40. Berlin: Akademie der Wissenschaften der DDR, Zentralinstitut für Sprachwissenschaft.

Ekström, Mats. 2000. “Information, Storytelling and Attractions: TV Journalism in Three Modes of Communication.” *Media, Culture & Society* 22 (4): 465–92. doi:10.1177/016344300022004006.

Elias, Norbert. 1978. *What is Sociology?* London: Hutchinson.

Engel, Ulrich. 1988. *Deutsche Grammatik.* Heidelberg: Julius Groos.

Ermida, Isabel. 2009. “When the Trails Go Cold: British Press Narratives on Maddie.” *Brno Studies in English* 35 (1): 45–62.

Ersson, Eugenia Olavarría de, and Philip Shaw. 2003. “Verb Complementation Patterns in Indian Standard English.” *English World-Wide* 24 (2): 137–61. doi:10.1075/eww.24.2.02ers.

Escorcia, Blanca de. 1985. “ESP and Beyond: A Quest for Relevance.” In *English in the World: Teaching and Learning the Language and Literatures; Papers of an International Conference Entitled 'Progress in English Studies' Held in London, 17 - 21 September 1984 to Celebrate the Fiftieth Anniversary of the British Council and Its Contribution to the Field of English Studies over Fifty Years*, edited by Randolph Quirk and Henry G. Widdowson, 228–37. Cambridge: Cambridge University Press.

Esser, Jürgen. 2009. *Introduction to English Text-Linguistics.* Textbooks in English Language and Linguistics (TELL) 2. Frankfurt am Main, New York: Peter Lang.

Estling, Maria. 2000. “Competition in the Wastebasket: A Study of Constructions with "All", "Both" and "Half".” In *Corpus Linguistics and Linguistic Theory*, edited by Christian Mair and Marianne Hundt, 103–16. Language and Computers: Studies in Practical Linguistics 33. Amsterdam: Rodopi.

Ethnologue. 2016. “Ethnologue: Languages of the World.” Accessed January 13, 2016. www.ethnologue.com.

Evans, Stephen. 2009. “The Evolution of the English-Language Speech Community in Hong Kong.” *English World-Wide* 30 (3): 278–301.

Evans, Stephen, and Christopher Green. 2001. “Language in Post-Colonial Hong Kong: The Roles of English and Chinese in the Public and Private Sectors.” *English World-Wide* 22 (2): 247–68.

Evert, Stefan, and The OCWB Development Team. 2010. “The IMS Open Corpus Workbench (CWB) CQP Query Language Tutorial.” http://cwb.sourceforge.net/files/CQP_Tutorial/.

Fahmy, Shahira, James D. Kelly, and Yung S. Kim. 2007. “What Katrina Revealed: A Visual Analysis of the Hurricane Coverage by News Wires and U.S. Newspapers.” *Journalism & Mass Communication Quarterly* 84 (3): 546–61. doi:10.1177/107769900708400309.

Fairbank, John King. 1992. *China: a New History.* Cambridge, Massachusetts: Belknap Press of Harvard University Press.

Fairclough, Norman. 1989. *Language and Power.* Language in Social Life. London: Longman.

———. 1995. *Critical Discourse Analysis: The Critical Study of Language.* Language in Social Life. London: Longman.

Ferguson, Charles A. 1983. “Sports Announcer Talk: Syntactic Aspects of Register Variation.” *Language in Society* 12 (2): 153–72.

Ferguson, Gibson. 2001. “If You Pop over There: A Corpus-Based Study of Conditionals in Medical Discourse.” *English for Specific Purposes* 20 (1): 61–82. doi:10.1016/S0889-4906(99)00027-7.

Fest, Jennifer. Forthcoming. “Sports News around the World – A Quantitative Study of Register Features in Newspaper Sports Coverage.” In *The Discourse of Sport: Analyses from Social Linguistics*, edited by David Caldwell, John Walsh, Elaina W. Vine, and Jon Jureidini. Routledge Studies in Sociolinguistics. London, New York: Routledge.

———. 2011. “A Corpus-Based Analysis of Register Features in British Football Journalism.” MA-Thesis, Department of English, American and Romance Studies, RWTH Aachen University.

———. 2015. “Defining Endings in Newspaper Writing – A Case Study on Football Coverage.” In *Last Things: Essays on Ends and Endings*, edited by Gavin Hopps, Stella Neumann, Sven Strasen, and Peter Wenzel, 53–64. Aachen British and American Studies 19. Frankfurt am Main: Peter Lang.

———. 2016. “Thematic Diversity and Distance in Newspaper Lifestyle Articles.” Sociolinguistics Symposium 21, Universidad de Murcia, Spain, June 15.

Festinger, Leon. 1957. *A Theory of Cognitive Dissonance.* Stanford: Stanford University Press.

Financial Times. 2014. “Circulation.” Accessed January 26, 2016. http://www.ftasiaads.com/downloads/Circulation.pdf.

Firth, John R. 1949. “The Semantics of Linguistic Science.” *Lingua* 1: 393–404.

———. 1950. “Personality and Language in Society.” *The Sociological Review* a42 (1): 37–52.

———. 1969. *Papers in Linguistics 1934-1951.* London: Oxford University Press.

Fischer, Lorenz, and Günther Wiswede. 1997. *Grundlagen der Sozialpsychologie.* München: Oldenbourg.

Fishman, Joshua A. 1996. “Introduction: Some Empirical and Theoretical Issues.” In *Post-Imperial English: Status Change in Former British and American Colonies, 1940-1990*, edited by Joshua A. Fishman, Andrew W. Conrad, and Alma Rubal-Lopez, 3–12. New York: De Gruyter Mouton.

Fishman, Joshua A., Andrew W. Conrad, and Alma Rubal-Lopez, eds. 1996. *Post-Imperial English: Status Change in Former British and American Colonies, 1940-1990.* New York: De Gruyter Mouton.

Fitzmaurice, Susan M. 2002. *The Familiar Letter in Early Modern English: A Pragmatic Approach.* Pragmatics & Beyond New Series 95. Amsterdam: John Benjamins.

Flew, Terry. 2009. "Democracy, Participation and Convergent Media: Case Studies in Contemporary Online News Journalism in Australia." *Communication, Politics & Culture* 42 (2): 87–109.

Flower, Wayne. 2011. "Train accident victim Ethan a 'cheeky, affectionate little boy'." *Herald Sun*, October 6. Accessed November 10, 2016. http://www.heraldsun.com.au/news/train-accident-victim-ethan-a-cheeky-affectionate-little-boy/story-e6frf7l6-1226159606815.

Flowerdew, John, David C. S. Li, and Sarah Tran. 2002. "Discriminatory News Discourse: Some Hong Kong Data." *Discourse & Society* 13 (3): 319–45. doi:10.1177/0957926502013003052.

Flowerdew, John, and Tony Dudley-Evans. 2002. "Genre Analysis of Editorial Letters to International Journal Contributors." *Applied Linguistics* 23 (4): 463–89. doi:10.1093/applin/23.4.463.

Flowerdew, John, and Solomon Leong. 2010. "Presumed Knowledge in the Discursive Construction of Socio-Political and Cultural Identity." *Journal of Pragmatics* 42 (8): 2240–52. doi:10.1016/j.pragma.2010.01.007.

Ford, Christopher A. 2010. *The Mind of Empire: China's History and Modern Foreign Relations.* Lexington: University Press of Kentucky.

Fowler, Roger. 1991. *Language in the News: Discourse and Ideology in the Press.* London: Routledge.

Fox, Annie B., Danuta Bukatko, Mark Hallahan, and Mary Crawford. 2007. "The Medium Makes a Difference: Gender Similarities and Differences in Instant Messaging." *Journal of Language and Social Psychology* 26 (4): 389–97. doi:10.1177/0261927X07306982.

Franklin, Bob. 1997. *Newszak and News Media.* London: Edward Arnold.

Franklin, Bob, Martin Hamer, Mark Hanna, Marie Kinsey, and John E. Richardson. 2005. *Key Concepts in Journalism Studies.* Sage Key Concepts. London: Sage Publications.

Friedlander, Edward J. 1982. *A Concise Guide to Newspaper Feature Writing.* Washington D.C. University Press of America.

Fries, Udo, and Peter Schneider. 2000. "ZEN: Preparing the Zurich English Newspaper Corpus." In *English Media Texts - Past and Present*, edited by Friedrich Ungerer, 3–24. Pragmatics & Beyond New Series 80. Amsterdam: John Benjamins.

Fryer, Daniel L. 2013. "Exploring the Dialogism of Academic Discourse: Heteroglossic Engagement in Medical Research Articles." In *English Corpus Linguistics: Variation in Time, Space and Genre*, edited by Gisle Andersen and Kristin Bech, 183–207. Language and Computers: Studies in Practical Linguistics 77. Amsterdam: Rodopi.

Fu, King-Wa, Yuen-Ying Chan, and Paul S. F. Yip. 2011. "Newspaper Reporting of Suicides in Hong Kong, Taiwan and Guangzhou: Compliance with WHO Media Guidelines and Epidemiological Comparisons." *Journal of Epidemiology and Community Health* 65 (10): 928–33. doi:10.1136/jech.2009.105650.

Fung, Anthony Y. H. 2007. "Political Economy of Hong Kong Media: Producing a Hegemonic Voice." *Asian Journal of Communication* 17 (2): 159–71. doi:10.1080/01292980701306530.

Fuoli, Matteo. 2013. "Texturing a Responsible Corporate Identity: A Comparative Analysis of Appraisal in BP'S and IKEA'S 2009 Corporate Social Reports." In *English Corpus Linguistics: Variation in Time, Space and Genre*, edited by Gisle Andersen and Kristin Bech, 209–35. Language and Computers: Studies in Practical Linguistics 77. Amsterdam: Rodopi.

Fuoli, Matteo, and Carita Paradis. 2014. "A Model of Trust-Repair Discourse." *Journal of Pragmatics* 74:52–69. doi:10.1016/j.pragma.2014.09.001.

Gabrielatos, Costas, and Paul Baker. 2008. "Fleeing, Sneaking, Flooding: A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press, 1996-2005." *Journal of English Linguistics* 36 (1): 5–38. doi:10.1177/0075424207311247.

García, Begoña C. 2004. "The Scientific Register in the History of English: A Corpus-Based Study." *Studia Neophilologica* 76 (2): 125–39. doi:10.1080/00393270410003611.

Geisler, Christer. 2002. "Investigating Register Variation in Nineteenth-Century English: A Multi-Dimensional Comparison." In *Using Corpora to Explore Linguistic Variation*, edited by Randi Reppen, Susan M. Fitzmaurice, and Douglas Biber, 249–71. Studies in Corpus Linguistics 9. Amsterdam: John Benjamins.

Gentzkow, Matthew, and Jesse Shapiro. 2006. *What Drives Media Slant? Evidence from U.S. Daily Newspapers.* Cambridge, MA: National Bureau of Economic Research.

Ghadessy, Mohsen, ed. 1988. *Registers of Written English: Situational Factors and Linguistic Features.* London: Pinter Publishers.

———. 1988. "The Language of Written Sports Commentary: Soccer: A Description." In *Registers of Written English: Situational Factors and Linguistic Features*, edited by Mohsen Ghadessy, 17–51. London: Pinter Publishers.

Gillen, Julia, and Guy Merchant. 2013. "Contact Calls: Twitter as a Dialogic Social and Linguistic Practice." *Language Sciences* 35:47–58. doi:10.1016/j.langsci.2012.04.015.

Gilles, Roger. 1997. "The Rhetoric of News and the Teaching of Writing." *Michigan Academician: Papers of the Michigan Academy of Science, Arts, and Literature* 29 (1): 69–77.

Gilquin, Gaëtanelle, and Sylviane Granger. 2011. "From EFL to ESL: Evidence from the International Corpus of Learner English." In *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*, edited by Joybrato Mukherjee and Marianne Hundt, 55–78. Amsterdam: John Benjamins.

Gimenez, Julio C. 2000. "Business E-Mail Communication: Some Emerging Tendencies in Register." *English for Specific Purposes* 19 (3): 237–51. doi:10.1016/S0889-4906(98)00030-1.

Gisborne, Nikolas. 2002. "Relative Clauses in Hong Kong English." In *Hong Kong English: Autonomy and Creativity*, edited by Kingsley Bolton, 141–60. Aberdeen: Hong Kong University Press.

———. 2009. "Aspects of the Morphosyntactic Typology of Hong Kong English." *English World-Wide* 30 (2): 149–69. doi:10.1075/eww.30.2.03gis.

Glass, Tom. 2009. "Why Thais Write to Other Thais in English." *World Englishes* 28 (4): 532–43. doi:10.1111/j.1467-971X.2009.01610.x.

Gledhill, Chris. 2000. "The Discourse Function of Collocation in Research Article Introductions." *English for Specific Purposes* 19 (2): 115–35. doi:10.1016/S0889-4906(98)00015-5.

Goldwert, Lindsay. 2012. "Men obsessed with their muscles are more sexist, more likely to believe thin women are more attractive." *Daily News*, June 26. Accessed November 10, 2016. http://www.nydailynews.com/life-style/men-obsessed-muscles-sexist-thin-women-attractive-article-1.1102522.

Görlach, Manfred. 1989. "Word-Formation and the ENL: ESL: EFL Distinction." *English World-Wide* 10 (2): 279–313.

———. 1990. *Studies in the History of the English Language.* Heidelberg: Winter.

———. 1996. "And is it English?" *English World-Wide* 17 (2): 153–74. doi:10.1075/eww.17.2.02gor.

———. 2012. "Variation in Academic Texts." In *Academic Identity Traits: A Corpus-Based Investigation*, edited by Maurizio Gotti, 23–42. Linguistic Insights: Studies in Language and Communication 150. Bern: Peter Lang.

Götz, Sandra, and Marco Schilk. 2011. "Formulaic Sequences in Spoken ENL, ESL and EFL: Focus on British English, Indian English and Learner English of Advanced German Learners." In *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*, edited by Joybrato Mukherjee and Marianne Hundt, 79–100. Amsterdam: John Benjamins.

Granato, Len. 2002. *Newspaper Feature Writing.* Sydney: University of New South Wales Press in Association with Deakin University Press.

Gray, Jack. 2002. *Rebellions and Revolutions: China from the 1800s to 2000.* 2nd ed. Short Oxford History of the Modern World. Oxford: Oxford University Press.

Greenbaum, Sidney. 1988. "A Proposal for an International Computerised Corpus of English." *World Englishes* 7 (3): 315. doi:10.1111/j.1467-971X.1988.tb00241.x.

———. 1990. "Standard English and the International Corpus of English." *World Englishes* 9 (1): 79–83. doi:10.1111/j.1467-971X.1990.tb00688.x.

Greenberg, David, and Galina Miazhevich. 2012. "Assimilationism vs. Multiculturalism: U.S. Identity and Media Representations of British Muslims over a 12-Year Span." *Communication, Culture & Critique* 5 (1): 75–98. doi:10.1111/j.1753-9137.2011.01117.x.

Greimas, Algirdas J. (1966) 1983. *Sémantique Structurale*. Paris: Larousse. Translated by McDowell, Daniele; Ronald Schleifer, and Alan Velie as *Structural Semantics: An Attempt at a Method* (Nebraska: University of Nebraska Press, 1983).

Grice, H. P. 1979a. "Intendieren, Meinen, Bedeuten." In *Handlung, Kommunikation, Bedeutung*, edited by Georg Meggle, 2–15. Frankfurt am Main: Suhrkamp.

———. 1979b. "Logik und Konversation." In *Handlung, Kommunikation, Bedeutung*, edited by Georg Meggle, 243–65. Frankfurt am Main: Suhrkamp.

Gries, Stefan T., and Joybrato Mukherjee. 2010. "Lexical Gravity across Varieties of English: an ICE-Based Study of N-Grams in Asian Englishes." *International Journal of Corpus Linguistics* 15 (4): 520–48. doi:10.1075/ijcl.15.4.04gri.

Gries, Stefan T., John Newman, and Cyrus Shaoul. 2011. "N-Grams and the Clustering of Registers." *Empirical Language Research Journal* 5 (1). http://ejournals.org.uk/ELR/article/2011/1. Accessed January 02, 2016.

Gross, Alan G., Joseph E. Harmon, and Michael S. Reidy. 2002. *Communicating Science: The Scientific Article from the 17th Century to the Present.* Oxford: Oxford University Press.

Gülich, Elisabeth, and Wolfgang Raible. 1977. *Linguistische Textmodelle: Grundlagen und Möglichkeiten.* München: Wilhelm Fink Verlag.

Gunnemark, Erik, and Donald Kenrick. 1983. *What Language Do They Speak?* Kinna, Sweden: Gunnemark.

Gupta, Anthea F. 1994. *The Step-Tongue: Children's English in Singapore.* Clevedon: Multilingual Matters.

Gut, Ulrike. 2011. "Studying Structural Innovations in New Englishes." In *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*, edited by Joybrato Mukherjee and Marianne Hundt, 101–24. Amsterdam: John Benjamins.

Guz, Wojciech. 2009. “English Affixal Nominalizations across Language Registers.” *Poznań Studies in Contemporary Linguistics* 45 (4): 461–85. doi:10.2478/v10010-009-0030-6.

Haase, Christoph. 2004. “Conceptualization Specifics in East African English: Quantitative Arguments from the ICE-East Africa Corpus.” *World Englishes* 23 (2): 261–68. doi:10.1111/j.0883-2919.2004.00350.x.

Haase, Fee-Alexandra. 2013. “A Contrastive Study of Essay Writing in German and English and Its Tradition of Rhetoric: The School Essay, Essay as Academic Writing, and the Essay as an Artificial Literary Product.” *Linguistica e Letteratura* 38 (1-2): 299–321.

Hagen, Ingunn. 1994. “The Ambivalences of TV News Viewing: Between Ideals and Everyday Practices.” *European Journal of Communication* 9 (2): 193–220. doi:10.1177/0267323194009002005.

Halliday, Michael A. K. 1961. “Categories of the Theory of Grammar.” *Word* 17 (3): 241–92.

———. 1973. *Explorations in the Functions of Language.* London: Edward Arnold.

———. 1974. *Language and Social Man.* London: Longman.

———. 1975. *Learning How to Mean - Explorations in the Development of Language.* London: Edward Arnold.

———. 1978. *Language as a Social Semiotic: The Social Interpretation of Language and Meaning.* London: Edward Arnold.

———. 1994. “Spoken and Written Modes of Meaning.” In *Media Texts: Authors and Readers*, edited by David Graddol and Oliver Boyd-Barrett, 51–73. Clevedon: Multilingual Matters.

———. 2001. “Literacy and Linguistics: Relationships between Spoken and Written Language.” In *Analysing English in a Global Context*, edited by Anne Burns and Caroline Coffin, 181–93. London, England: Routledge, with Macquarie University and Open University.

———. 2006. “Systemic Theory.” In *Encyclopedia of Language and Linguistics*. Vol. 12, edited by Keith Brown. 14 vols, 443–48. Amsterdam: Elsevier.

Halliday, Michael A. K., and Ruqaiya Hasan. 1976. *Cohesion in English.* London: Longman.

———. 1985. *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective.* Oxford: Oxford University Press.

Halliday, Michael A. K., and James R. Martin. 1993. *Writing Science: Literacy and Discursive Power.* Pittsburgh: University of Pittsburgh Press.

Halliday, Michael A. K., and Christian M. I. M. Matthiessen. 1999. *Construing Experience through Meaning: A Language-Based Approach to Cognition.* Open Linguistics Series. London: Cassell.

———. 2014. *An Introduction to Functional Grammar.* 4th ed. London: Edward Arnold.

Halliday, Michael A. K., Angus McIntosh, and Peter Strevens. 1964. *The Linguistic Sciences and Language Teaching.* London: Longman.

Hamilton, J. W. 1928. *World or Cosmo English: a Proposal to Simplify English as the International Auxiliary Language for the World:* private publishing.

Hancock, Ian F., and Rachel Angogo. 1986. “English in East Africa.” In *English as a World Language*, edited by Richard W. Bailey and Manfred Görlach. 4th ed., 306–23. Ann Arbor: The University of Michigan Press.

Hänsel, Eva C., and Dagmar Deuber. 2013. “Globalization, Postcolonial Englishes, and the English Language Press in Kenya, Singapore, and Trinidad and Tobago.” *World Englishes* 32 (3): 338–57. doi:10.1111/weng.12035.

Hansen-Ampah, Adjan. 2013. *Header Tool.* Aachen.

Harcup, Tony. 2005. *Journalism: Principles and Practice.* Reprint. London: Sage Publications.

Harcup, Tony, and Deirdre O'Neill. 2001. "What is News? Galtung and Ruge Revisited." *Journalism Studies* 2 (2): 261–80. doi:10.1080/14616700118449.

Hardy, Jack, and Eric Friginal. 2012. "Filipino and American Online Communication and Linguistic Variation." *World Englishes* 31 (2): 143–61. doi:10.1111/j.1467-971X.2011.01728.x.

———. 1987. "Diskussionstexte: Argumente für eine Systembetrachtung der Textorganisation." In *Sprache und Pragmatik*, edited by Inger Rosengren, 7–31. Stockholm: Lunder Symposium 1986.

Hart, Simon. 2013. "Heather Watson remains hopeful of making French Open in May after being laid low by glandular fever." *The Daily Telegraph*, April 10. Accessed November 10, 2016. http://www.telegraph.co.uk/sport/tennis/9985120/Heather-Watson-remains-hopeful-of-making-French-Open-in-May-after-being-laid-low-by-glandular-fever.html.

Harweg, Roland. 1968. *Pronomina und Textkonstitution.* München: Fink.

Hasan, Ruqaiya. 1978. "Text in the Systemic-Functional Model." In *Current Trends in Textlinguistics*, edited by Wolfgang Dressler, 228–46. Research in Text Theory 2. Berlin: De Gruyter.

———. 1999. "Speaking with Reference to Context." In *Text and Context in Functional Linguistics*, edited by Mohsen Ghadessy, 219–328. Current Issues in Linguistic Theory 169. Amsterdam: John Benjamins.

———. 2009a. "A View of Pragmatics in a Social Semiotic Perspective." *Linguistics and the Human Sciences* 5 (3): 251–79. doi:10.1558/lhs.v5i3.251.

———. 2009b. "The Place of Context in a Systemic Functional Model." In *Continuum Companion to Systemic Functional Linguistics*, edited by Michael A. K. Halliday and Jonathan Webster, 166–89. London: Continuum.

Hassan, Riaz. 1995. "Effects of Newspaper Stories on the Incidence of Suicide in Australia: A Research Note." *Australian and New Zealand Journal of Psychiatry* 29 (3): 480–83. doi:10.3109/00048679509064957.

Hawes, Thomas, and Sarah Thomas. 1995. "Language Bias against Women in British and Malaysian Newspapers." *Australian Review of Applied Linguistics* 18 (2): 1–18.

Haynes, Lilith. 1983. "Caribbean English: Form and Function." In *The Other Tongue: English across Cultures*, edited by Braj B. Kachru. 1st ed., 210–26. Oxford: Pergamon Press.

He, Deyuan, and David C. S. Li. 2009. "Language Attitudes and Linguistic Features in the 'China English' Debate." *World Englishes* 28 (1): 70–89. doi:10.1111/j.1467-971X.2008.01570.x.

Hecklau, Hans. 1977. "Struktur, Entwicklung und Verteilung der Bevölkerung." In *Kenia: Geographie, Vorgeschichte, Geschichte, Gesellschaft, Kultur, Erziehung, Gesundheitswesen, Wirtschaft, Entwicklung*, edited by Walter Leifer, 179–94. Tübingen: Horst Erdmann Verlag.

Heine, Bernd. 1977. "Sprachen und Sprachprobleme in Kenia." In *Kenia: Geographie, Vorgeschichte, Geschichte, Gesellschaft, Kultur, Erziehung, Gesundheitswesen, Wirtschaft, Entwicklung*, edited by Walter Leifer, 251–67. Tübingen: Horst Erdmann Verlag.

Heinemann, Margot, and Wolfgang Heinemann. 2002. *Grundlagen der Textlinguistik: Interaktion - Text - Diskurs.* Reihe Germanistische Linguistik 230. Tübingen: Niemeyer.

Hendry, Sharon. 2007. "Stocks, shocks and crashes." *The Sun*, October 4.

Henningham, John. 1996. "The Shape of Daily News: a Content Analysis of Australia's Metropolitan Newspapers." *Media International Australia* 79 (1): 22–34. doi:10.1177/1329878X9607900104.

Henry, Alex, and Robert L. Roseberry. 2001. "A Narrow-Angled Corpus Analysis of Moves and Strategies of the Genre: 'Letter of Application'." *English for Specific Purposes* 20 (2): 153–67. doi:10.1016/S0889-4906(99)00037-X.

Hepp, Andreas. 2004. *Netzwerke der Medien: Medienkulturen und Globalisierung.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Hermida, Alfred, and Neil Thurman. 2008. "A Clash of Cultures: The Integration of User-Generated Content within Professional Journalistic Frameworks at British Newspaper Websites." *Journalism Practice* 2 (3): 343–56. doi:10.1080/17512780802054538.

Herring, Susan C., and John C. Paolillo. 2006. "Gender and Genre Variation in Weblogs." *Journal of Sociolinguistics* 10 (4): 439–59. doi:10.1111/j.1467-9841.2006.00287.x.

Hiraga, Yuko. 2005. "British Attitudes towards Six Varieties of English in the USA and Britain." *World Englishes* 24 (3): 289–308. doi:10.1111/j.0883-2919.2005.00411.x.

Hoey, Michael. 1991. *Patterns of Lexis in Text: Describing English language.* Oxford: Oxford University Press.

Holland, Patricia. 1998. "The Politics of the Smile: 'Soft News' and the Sexualisation of the Popular Press." In *News, Gender, and Power*, edited by Cynthia Carter, Gill Branston, and Stuart Allan, 17–32. London: Routledge.

Holmes, Richard. 1997. "Genre Analysis, and the Social Sciences: An Investigation of the Structure of Research Article Discussion Sections in Three Disciplines." *English for Specific Purposes* 16 (4): 321–37. doi:10.1016/S0889-4906(96)00038-5.

Holzknecht, Susanne. 1989. "Sociolinguistic Analysis of a Register: Birthday Notices in Papua New Guinea Post Courier." *World Englishes* 8 (2): 179–92. doi:10.1111/j.1467-971X.1989.tb00653.x.

Hong Kong Census and Statistics Department. 1982. "Preliminary Results of the 1981 Population Census." Accessed May 31, 2015. http://www.statistics.gov.hk/pub/B78201FA1982XXXXE0100.pdf.

———. 2012. "Thematic Household Survey - Report No. 49." Accessed November 08, 2015. http://www.statistics.gov.hk/pub/B11302492012XXXXB0100.pdf.

———. 2013. "Thematic Household Survey - Report No. 51." Accessed May 31, 2015. http://www.statistics.gov.hk/pub/B11302512013XXXXB0100.pdf.

———. 2015. "Hong Kong Monthly Digest of Statistics May 2015." Accessed May 27, 2015. http://www.statistics.gov.hk/pub/B10100022015MM05B0100.pdf.

Hong Kong Department of Justice. 2015. "The Basic Law." Accessed May 31, 2015. http://www.basiclaw.gov.hk/en/basiclawtext/images/basiclaw_full_text_en.pdf.

Hopkins, Stephen. 2011. "Yesterday's News? Journalists, the Northern Ireland 'Troubles' and the Politics of Memoir-Writing." *Nordic Irish Studies* 10: 97–118.

Horkheimer, Max, and Theodor W. Adorno. 1947. "Kulturindustrie: Aufklärung als Massenbetrug." In *Dialektik der Aufklärung: Philosophische Fragmente*, edited by Max Horkheimer and Theodor W. Adorno, 128–76. Amsterdam: Querido.

Huang, Kailai. 2001. "American Business and the China Trade Embargo in the 1950s." *Essays in Economic & Business History* 19: 33–48.

Hudson-Ettle, Diana, and Josef J. Schmied. 1999. "Manual to Accompany the East African Component of the International Corpus of English: Background Information, Coding Conventions and Lists of Source Texts." Accessed January 16, 2016. https://www.tu-chemnitz.de/phil/english/sections/linguist/real/independent/ICE-EA/.

Hundt, Marianne, and Katrin Vogel. 2011. "Overuse of the Progressive in ESL and Learner Englishes – Fact or Fiction?" In *Exploring Second-Language Varieties of English and Learner*

Englishes: Bridging a Paradigm Gap, edited by Joybrato Mukherjee and Marianne Hundt, 145–66. Amsterdam: John Benjamins.

Hunston, Susan. 2002. *Corpora in Applied Linguistics.* Cambridge: Cambridge University Press.

Hurlimann, Anna, and Sara Dolnicar. 2012. “Newspaper Coverage of Water Issues in Australia.” *Water Research* 46 (19): 6497–6507. doi:10.1016/j.watres.2012.09.028.

Hützen, Nicole. 2016. “Shell Nouns in Academic Journal Articles: Differences and Similarities across Eight Disciplines.” PhD, Department of English, American and Romance Studies, RWTH Aachen University.

Hyland, Ken. 1997. “Language Attitudes at the Handover: Communication and Identity in 1997 Hong Kong.” *English World-Wide* 18 (2): 191–210.

———. 1998. *Hedging in Scientific Research Articles.* Amsterdam: John Benjamins.

———. 2002. “Authority and Invisibility.” *Journal of Pragmatics* 34 (8): 1091–1112. doi:10.1016/S0378-2166(02)00035-8.

———. 2004. “Graduates' Gratitude: The Generic Structure of Dissertation Acknowledgements.” *English for Specific Purposes* 23 (3): 303–24. doi:10.1016/S0889-4906(03)00051-6.

Hymes, Dell. 1967. “Models of the Interaction of Language and Social Setting.” *Journal of Social Issues* 23 (2): 8–28. doi:10.1111/j.1540-4560.1967.tb00572.x.

———. 1974. *Foundations in Sociolinguistics: An Ethnographic Approach.* Philadelphia: University of Pennsylvania Press.

ICE Project. 2015. “International Corpus of English.” Accessed January 13, 2016. http://ice-corpora.net/ICE/INDEX.HTM.

Iedema, Rick. 1997. “The Structure of the Accident News Story.” *Australian Review of Applied Linguistics* 20 (2): 95–118.

International Federation of Audit Bureaux of Circulations. 2013. “Data Reports.” Accessed November 06, 2015. http://www.ifabc.org/resources/data-reports%2065/data-reports-65.

Isaacs, Victor, and Rod Kirkpatrick. 2003. *Two Hundred Years of Sydney Newspapers: a Short History.* North Richmond: Rural Press Ltd.

Isenberg, Horst. 1968. *Der Begriff "Text" in der Sprachtheorie.* Berichte der Arbeitsgruppe Strukturelle Grammatik 8. Berlin: Deutsche Akademie der Wissenschaft.

———. 1971. “Überlegungen zur Texttheorie.” In *Literaturwissenschaft und Linguistik*. Vol. 1, edited by Jens Ihwe. 3 vols, 155–72. Frankfurt am Main: Athenäum.

Ismail, Jamal A., and James Deane. 2008. “The 2007 General Election in Kenya and Its Aftermath: The Role of Local Language Media.” *The International Journal of Press/Politics* 13 (3): 319–27. doi:10.1177/1940161208319510.

Jabour, Bridie. 2012. “Road unlikely to have caused horrific crash, says mayor.” *The Sydney Morning Herald*, October 1. Accessed November 10, 2016. http://www.smh.com.au/queensland/road-unlikely-to-have-caused-horrific-crash-says-mayor-20121001-26u68.html.

Jaccarino, Mike. 2012. “How Instagram bet the farm on Facebook stock… and lost $300m when its shares tanked.” *Daily Mail*, August 21. Accessed November 05, 2016. http://www.dailymail.co.uk/news/article-2191695/How-Instagram-bet-farm-Facebook-stock--lost-300-MILLION-shares-tanked.html.

Jacobson, Arne. 2007. “Connective Power: Solar Electrification and Social Change in Kenya.” *World Development* 35 (1): 144–62. doi:10.1016/j.worlddev.2006.10.001.

Jakobson, Roman. 1960. “Closing Statement: Linguistics and Poetics.” In *Style in Language*, edited by Thomas A. Seboek, 350–77. Cambridge, MA: MIT Press.

Janda, Richard D. 1985. “Note-Taking English as a Simplified Register.” *Discourse Processes* 8 (4): 437–54. doi:10.1080/01638538509544626.

Jaworska, Sylvia, and Ramesh Krishnamurthy. 2012. “On the F Word: A Corpus-Based Analysis of the Media Representation of Feminism in British and German Press Discourse, 1990-2009.” *Discourse & Society* 23 (4): 401–31. doi:10.1177/0957926512441113.

Jeffery, Chris, and Bertus van Rooy. 2004. “Emphasizer 'now' in Colloquial South African English.” *World Englishes* 23 (2): 269–80. doi:10.1111/j.0883-2919.2003.00351.x.

Jensen, Astrid. 2009. “Discourse Strategies in Professional E-Mail Negotiation: A Case Study.” *English for Specific Purposes* 28 (1): 4–18. doi:10.1016/j.esp.2008.10.002.

Johnson, Harry M. 1966. *Sociology: A Systematic Introduction.* New Delhi: Allied Publishers.

Johnson, Robert K. 1994. “Language Policy and Planning in Hong Kong.” *Annual Review of Applied Linguistics* 14:177. doi:10.1017/S0267190500002889.

Johnson, Sally, and Astrid Ensslin. 2013. “'But Her Language Skills Shifted the Family Dynamics Dramatically’: Language, Gender and the Construction of Publics in Two British Newspapers.” *Gender and Language* 1 (2): 229–54. doi:10.1558/genl.v1.i1.17192.

Joseph, John. 2004. *Language and Identity: National, Ethnic, Religious.* Houndmills: Palgrave Macmillan.

Jucker, Andreas H. 1992. *Social Stylistics: Syntactic Variation in British Newspapers.* Berlin, New York: De Gruyter Mouton.

———. 1996. “News Actor Labelling in British Newspapers.” *Text - Interdisciplinary Journal for the Study of Discourse* 16 (3): 373–90. doi:10.1515/text.1.1996.16.3.373.

Juma, Victor. 2013. “Top Kenol owners transfer Sh1.2bn shares in months.” *Daily Nation*, October 13. Accessed November 10, 2016. http://www.nation.co.ke/business/Top-Kenol-owners-transfer-Sh1-2bn-shares-in-months-/-/996/2030104/-/8omt8p/-/index.html.

Kachru, Braj B. 1981. “The Pragmatics of Non-Native Varieties of English.” In *English for Cross-Cultural Communication*, edited by Larry E. Smith, 15–39. London: Palgrave Macmillan.

———. 1983a. “Introduction: The Other Side of English.” In *The Other Tongue: English across Cultures*, edited by Braj B. Kachru. 1st ed., 1–12. Oxford: Pergamon Press.

———. 1983b. “Meaning in Deviation: Toward Understanding Non-Native English Texts.” In *The Other Tongue: English across Cultures*, edited by Braj B. Kachru. 1st ed., 325–50. Oxford: Pergamon Press.

———. 1983c. “Models for Non-Native Englishes.” In *The Other Tongue: English across Cultures*, edited by Braj B. Kachru. 1st ed., 31–57. Oxford: Pergamon Press.

———. 1983d. *The Indianization of English: The English Language in India.* Delhi: Oxford University Press.

———. 1985. “Standards, Codification and Sociolinguistic Realism: The English in the Outer Circle.” In *English in the World: Teaching and Learning the Language and Literatures; Papers of an International Conference Entitled 'Progress in English Studies' Held in London, 17 - 21 September 1984 to Celebrate the Fiftieth Anniversary of the British Council and Its Contribution to the Field of English Studies over Fifty Years*, edited by Randolph Quirk and Henry G. Widdowson, 11–30. Cambridge: Cambridge University Press.

———. 1988. “The Sacred Cows of English.” *English Today* 4 (4): 3–8. doi:10.1017/S0266078400000973.

———. 1990. *The Alchemy of English: The Spread, Functions, and Models of Non-Native Englishes.* Urbana and Chicago: University of Illinois Press.

———. 1991. “Liberation Linguistics and the Quirk Concern.” *English Today* 7 (1): 3–13. doi:10.1017/S026607840000523X.

———. 1992. “Teaching World Englishes.” In *The Other Tongue: English across Cultures*, edited by Braj B. Kachru. 2nd ed., 355–65. Urbana and Chicago: University of Illinois Press.

———. 1996. “World Englishes: Agony and Ecstasy.” *Journal of Aesthetic Education* 30 (2): 135–55.

———. 1997. “World Englishes and English-Using Communities.” *Annual Review of Applied Linguistics* (17): 66–87.

———. 2006. “English: World Englishes.” In *Encyclopedia of Language and Linguistics*. Vol. 4, edited by Keith Brown. 14 vols, 195–202. Amsterdam: Elsevier.

Kachru, Braj B., and Randolph Quirk. 1981. “Introduction.” In *English for Cross-Cultural Communication*, edited by Larry E. Smith, xiii–xx. London: Palgrave Macmillan.

Kachru, Braj B., and Larry E. Smith. 1985. “Editorial.” *World Englishes* 4 (2): 209–12. doi:10.1111/j.1467-971X.1985.tb00408.x.

Kachru, Yamuna, and Cecil L. Nelson. 2006. *World Englishes in Asian Contexts.* Asian Englishes today. Hong Kong: Hong Kong University Press.

Kallen, Jeffrey. 1991. “Sociolinguistic Variation and Methodology: 'After' as a Dublin Variable.” In *English around the World: Sociolinguistic Perspectives*, edited by Jenny Cheshire, 61–74. Cambridge: Cambridge University Press.

Kane, Mary J., and Helen J. Lenskyi. 1998. “Media Treatment of Female Athletes: Issues of Gender and Sexuality.” In *Mediasport*, edited by Lawrence A. Wenner, 186–201. London: Routledge.

Kanoksilapatham, Budsaba. 2005. “Rhetorical Structure of Biochemistry Research Articles.” *English for Specific Purposes* 24 (3): 269–92. doi:10.1016/j.esp.2004.08.003.

Kanyoro, Musimbi R. A. 1991. “The Politics of the English Language in Kenya and Tanzania.” In *English around the World: Sociolinguistic Perspectives*, edited by Jenny Cheshire, 402–19. Cambridge: Cambridge University Press.

Katzenberger, Stefan. 1999. *Komplementäre Kommunikation lokaler Medien: Hörfunk und Presse: intermediale Profile, programmliche Pools, publizistische Performanz.* Aktuelle Medien- und Kommunikationsforschung 9. Münster: LIT Verlag.

Keeble, Richard. 2006. *The Newspapers Handbook.* 4th ed. London: Routledge.

Kenya National Bureau of Statistics. 2010. “The 2009 Kenya Population and Housing Census: Volume II: Population and Household Distribution by Socio-Economic Characteristics.” Accessed February 05, 2016. http://www.knbs.or.ke/index.php?option=com_phocadownload&view=category&id=109:population-and-housing-census-2009&Itemid=599.

———. 2014. “Kenya Facts and Figures 2014.” Accessed November 08, 2015. http://www.knbs.or.ke/index.php?option=com_phocadownload&view=category&id=20&Itemid=595.

Kenya Presidency. 2016. “Official Government Website: Kenya in Brief.” Accessed January 26, 2016. http://presidency.go.ke/index.php/kenya-in-brief.

Kenyan National Council for Law Reporting with the Authority of the Attorney General. 2010. “The Constitution of Kenya.” Accessed May 23, 2015. https://www.kenyaembassy.com/pdfs/The%20Constitution%20of%20Kenya.pdf.

Kenyon, John S. 1924. *American Pronunciation.* Ann Arbor: George Wahr.

KhosraviNik, Majid. 2008. “British Newspapers and the Representation of Refugees, Asylum Seekers and Immigrants between 1996 and 2006.” *Centre for Language in Social Life Working Papers* 128. http://www.lancaster.ac.uk/fass/groups/clsl/docs/clsl128.pdf. Accessed January 09, 2016.

———. 2009. “The Representation of Refugees, Asylum Seekers and Immigrants in British Newspapers during the Balkan Conflict (1999) and the British General Election (2005).” *Discourse & Society* 20 (4): 477–98. doi:10.1177/0957926509104024.

———. 2010a. “Actor Descriptions, Action Attributions, and Argumentation: Towards a Systematization of CDA Analytical Categories in the Representation of Social Groups.” *Critical Discourse Studies* 7 (1): 55–72. doi:10.1080/17405900903453948.

———. 2010b. “The Representation of Refugees, Asylum Seekers and Immigrants in British Newspapers: A Critical Discourse Analysis.” *Journal of Language and Politics* 9 (1): 1–28. doi:10.1075/jlp.9.1.01kho.

Kim, Kyung H. 2014. “Examining US News Media Discourses about North Korea: A Corpus-Based Critical Discourse Analysis.” *Discourse & Society* 25 (2): 221–44. doi:10.1177/0957926513516043.

Kim, Yong-Jin, and Douglas Biber. 1994. “A Corpus-Based Analysis of Register Variation in Korean.” In *Sociolinguistic Perspectives on Register*, edited by Douglas Biber and Edward Finegan, 157–81. Oxford: Oxford University Press.

King, Cynthia, and Paul M. Lester. 2005. “Photographic Coverage during the Persian Gulf and Iraqi Wars in Three U.S. Newspapers.” *Journalism & Mass Communication Quarterly* 82 (3): 623–37. doi:10.1177/107769900508200309.

Kirkpatrick, Rod. 2001. “Are Community Newspapers Really Different?” *Asia Pacific Media Educator* 1 (10): 16–21.

Klapper, Joseph T. 1957. “What We Know About the Effects of Mass Communication: The Brink of Hope.” *Public Opinion Quarterly* 21 (4): 453–74. doi:10.1086/266744.

———. 1960. *The Effects of Mass Communication.* New York: Free Press.

Kleinke, Sonja. 2000. “Women and Headline-Policy in German and English Local Daily Newspapers.” In *English Media Texts - Past and Present*, edited by Friedrich Ungerer, 67–84. Pragmatics & Beyond New Series 80. Amsterdam: John Benjamins.

Kleinnijenhuis, Jan, Friederike Schultz, Sonja Utz, and Dirk Oegema. 2015. “The Mediating Role of the News in the BP Oil Spill Crisis 2010: How U.S. News Is Influenced by Public Relations and in Turn Influences Public Awareness, Foreign News, and the Share Price.” *Communication Research* 42 (3): 408–28. doi:10.1177/0093650213510940.

Klerk, Vivian de. 2005. “The Use of 'actually' in Spoken Xhosa English: A Corpus Study.” *World Englishes* 24 (3): 275–88. doi:10.1111/j.0883-2919.2005.00410.x.

Klocker, Natascha, and Kevin M. Dunn. 2003. “Who's Driving the Asylum Debate: Newspaper and Government Representations of Asylum Seekers.” *Media International Australia* 109: 71–92.

Knox, John S. 2009a. “Multimodal Discourse on Online Newspaper Home Pages: A Social-Semiotic Perspective.” PhD Thesis, Department of Linguistics, University of Sydney.

———. 2009b. “Punctuating the Home Page: Image as Language in an Online Newspaper.” *Discourse & Communication* 3 (2): 145–72. doi:10.1177/1750481309102450.

Koester, Almut J. 2006. *Investigating Workplace Discourse.* London: Routledge.

Kolodzy, Janet. 2006. *Convergence Journalism: Writing and Reporting across the News Media.* Lanham, Oxford: Rowman & Littlefield.

Kopple, William J. V. 1998. “Relative Clauses in Spectroscopic Articles in the Physical Review, Beginnings and 1980: Some Changes in Patterns of Modification and a Connection to a Possible Shift in Style.” *Written Communication* 15 (2): 170–202. doi:10.1177/0741088398015002002.

Kornetzki, Anastasiya. 2012. *Contrastive Analysis of News Text Types in Russian, British and American Business Online and Print Media.* Forum für Fachsprachen-Forschung 102. Berlin: Frank & Timme.

Kortmann, Bernd, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie, and Clive Upton, eds. 2004. *A Handbook of Varieties of English.* 2 vols. Berlin: De Gruyter Mouton.

Kortmann, Bernd, and Benedikt Szmrecsanyi. 2004. “Global Synopsis: Morphological and Syntactic Variation in English.” In *A Handbook of Varieties of English*. Vol. 2, edited by Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie, and Clive Upton. 2 vols, 1142–1202. Berlin: De Gruyter Mouton.

Koteyko, Nelya, Brian Brown, and Paul Crawford. 2008. “The Dead Parrot and the Dying Swan: The Role of Metaphor Scenarios in UK Press Coverage of Avian Flu in the UK in 2005–2006.” *Metaphor and Symbol* 23 (4): 242–61. doi:10.1080/10926480802426787.

Kress, Gunther. 1993. “Genre as Social Process.” In *The Powers of Literacy: A Genre Approach to Teaching Writing*, edited by Bill Cope and Mary Kalantzis, 22–37. Pittsburgh: University of Pittsburgh Press.

Kress, Gunther, and Robert Hodge. 1979. *Language as Ideology.* London: Routledge and Kegan Paul.

Krings, Constanze. 2003. *Zur Typologie des Erzählschlusses in der englischsprachigen Kurzgeschichte.* Aachen British and American Studies. Frankfurt am Main: Peter Lang.

Lamb, Kate. 2013. “Australian stabbed during Bali robbery.” *The Sydney Morning Herald*, January 4. Accessed November 10, 2016. http://www.smh.com.au/national/australian-stabbed-during-bali-robbery-20130104-2c8s7.html.

Lamb, Sharon. 1991. “Acts without Agents: An Analysis of Linguistic Avoidance in Journal Articles on Men Who Batter Women.” *American Journal of Orthopsychiatry* 61 (2): 250–57. doi:10.1037/h0079243.

Lanson, Jerry, and Barbara C. Fought. 1999. *News in a New Century: Reporting in an Age of Converging Media.* Thousand Oaks: Pine Forge Press.

Larcinese, Valentino, Riccardo Puglisi, and James M. Snyder. 2011. “Partisan Bias in Economic News: Evidence on the Agenda-Setting Behavior of U.S. Newspapers.” *Journal of Public Economics* 95 (9-10): 1178–89. doi:10.1016/j.jpubeco.2011.04.006.

Lau, Kimberley. 1996. “'It's About Time': The Ten Proverbs Most Frequently Used in Newspapers and Their Relation to American Values.” *Proverbium* 13: 135–59.

Lavid, Julia. 1995. “Towards a Text Type Taxonomy: A Functional Framework for Text Analysis and Generation.” *Procesamiento del lenguaje natural* 16: 29–43.

Lavric, Eva, Gerhard Pisek, Andrew C. Skinner, and Wolfgang Stadler, eds. 2008. *The Linguistics of Football.* Tübingen: Narr.

Lawton, David L. 1986. “English in the Caribbean.” In *English as a World Language*, edited by Richard W. Bailey and Manfred Görlach. 4th ed., 251–80. Ann Arbor: The University of Michigan Press.

Lazarsfeld, Paul F., Bernard Berelson, and Hazel Gaudet. 1944. *The People's Choice: How the Voter Makes up His Mind in a Presidential Campaign.* New York: Columbia University Press.

Lazarsfeld, Paul F., and Raymond Boudon. 1993. *On Social Research and Its Language.* Chicago, London: University of Chicago Press.

Lazarsfeld, Paul F., and Herbert Menzel. 1963. “Mass Media and Personal Influence.” In *The Science of Human Communication*, edited by Wilbur Schramm, 94–115. New York: Basic Books.

Lee, Chin-Chuan. 1998. “Press Self-Censorship and Political Transition in Hong Kong.” *The Harvard International Journal of Press/Politics* 3 (2): 55–73. doi:10.1177/1081180X98003002005.

———. 2000. “The Paradox of Political Economy: Media Structure, Press Freedom, and Regime Change in Hong Kong.” In *Power, Money, and Media: Communication Patterns and Bureaucratic Control in Cultural China*, edited by Chin-Chuan Lee, 288–336. Evanston: Northwestern University Press.

Lee, Chin-Chuan, Joseph M. Chan, Zhongdang Pan, and Clement So. 2002. *Global Media Spectacle: News War over Hong Kong.* SUNY Series in Global Media Studies. Albany, NY: State University of New York Press.

Lee, David. 2001. “Genres, Registers, Text Types, Domains and Styles: Clarifying the Concepts and Navigating a Path through the BNC Jungle.” *Language Learning and Technology* 5 (3): 37–72.

Lee, Francis L. 2004. “Constructing Perfect Women: The Portrayal of Female Officials in Hong Kong Newspapers.” *Media, Culture & Society* 26 (2): 207–25. doi:10.1177/0163443704041173.

Lee, Francis L., and Joseph M. Chan. 2008. “Organizational Production of Self-Censorship in the Hong Kong Media.” *The International Journal of Press/Politics* 14 (1): 112–33. doi:10.1177/1940161208326598.

Lee, Francis L., and Angel Lin. 2006. “Newspaper Editorial Discourse and the Politics of Self-Censorship in Hong Kong.” *Discourse & Society* 17 (3): 331–58. doi:10.1177/0957926506062371.

Lee, Jackie F. K., and Peter Collins. 2004. “On the Usage of 'have', 'dare', 'need', 'ought' and 'used To' in Australian English and Hong Kong English.” *World Englishes* 23 (4): 501–13. doi:10.1111/j.0083-2919.2004.00374.x.

Lee, Junghi, and Robert L. Craig. 1992. “News as an Ideological Framework: Comparing US Newspapers' Coverage of Labor Strikes in South Korea and Poland.” *Discourse & Society* 3 (3): 341–63. doi:10.1177/0957926592003003004.

Lee, Liv. 2013. “Murray ducks out of Great Britain's Davis Cup tie with Russia.” *Daily Mail*, January 31. Accessed November 09, 2016. http://www.dailymail.co.uk/sport/tennis/article-2271290/Andy-Murray-absent-Britains-Davis-Cup-tie-Russia.html#axzz2KEO5Gpa6.

Lee, Paul, and L. L. Chu. 1998. “Inherent Dependence on Power: The Hong Kong Press in Political Transition.” *Media, Culture & Society* 20 (1): 59–77. doi:10.1177/016344398020001005.

Lee, Wei-Na, and David K. Tse. 1994. “Changing Media Consumption in a New Home: Acculturation Patterns Among Hong Kong Immigrants to Canada.” *Journal of Advertising* 23 (1): 57–70. doi:10.1080/00913367.1994.10673431.

Leech, Geoffrey. 1991. “The State of the Art in Corpus Linguistics.” In *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, edited by Karin Aijmer and Bengt Altenberg, 105–22. London: Longman.

Leifer, Walter. 1977. “Von den Uranfängen bis zur Gegenwart.” In *Kenia: Geographie, Vorgeschichte, Geschichte, Gesellschaft, Kultur, Erziehung, Gesundheitswesen, Wirtschaft, Entwicklung*, edited by Walter Leifer, 103–61. Tübingen: Horst Erdmann Verlag.

Leith, William. 2012. "The bitter truth about sugar." *The Daily Telegraph*, March 27. Accessed November 10, 2016. http://www.telegraph.co.uk/health/dietandfitness/9160114/The-bitter-truth-about-sugar.html.

Leitner, Gerhard. 2000. "Lexical Frequencies in a 300 Million Word Corpus of Australian Newspapers. Analysis and Interpretation." *International Journal of Corpus Linguistics* 5 (2): 147–78. doi:10.1075/ijcl.5.2.04lei.

Leontovich, Olga A. 2005. "American English as a Medium of Intercultural Communication." *World Englishes* 24 (4): 523–32. doi:10.1111/j.0883-2919.2005.00436.x.

Leschke, Rainer. 2003. *Einführung in die Medientheorie.* München: Fink.

Lewis, Diana M. 2003. "Online News: A New Genre?" In *New Media Language*, edited by Jean Aitchison and Diana M. Lewis, 95–104. London: Routledge.

Lewis, Frank. 2013. "Sackey's en route." *The Sun*, April 8. Accessed November 06, 2016. http://www.thesun.co.uk/sol/homepage/sport/rugby_union/4878852/Paul-Sackey-en-route-back-to-England-after-French-exile.html.

Lewis, M. P., Gary F. Simons, and Charles D. Fennig. 2015. *Ethnologue: Languages of Kenya.* Dallas: SIL International. Accessed May 31, 2015.

LexisNexis. 2016. "LexisNexis Academic." Accessed May 09, 2016. http://www.lexisnexis.com/hottopics/lnacademic.

Li, David C. S. 2000. "Cantonese-English Code-Switching Research in Hong Kong: A Y2K Review." *World Englishes* 19 (3): 305–22. doi:10.1111/1467-971X.00181.

Li, Juan. 2011. "Collision of Language in News Discourse: A Functional-Cognitive Perspective on Transitivity." *Critical Discourse Studies* 8 (3): 203–19. doi:10.1080/17405904.2011.586231.

Li, Mu, Simon Chapman, Kingsley Agho, and Creswell J. Eastman. 2008. "Can Even Minimal News Coverage Influence Consumer Health-Related Behaviour? A Case Study of Iodized Salt Sales, Australia." *Health Education Research* 23 (3): 543–48. doi:10.1093/her/cym028.

Lim, Lisa, and Nikolas Gisborne. 2009. "The Typology of Asian Englishes: Setting the Agenda." *English World-Wide* 30 (2): 123–32. doi:10.1075/eww.30.2.01lim.

Lin, Angel, and Champa Detaramani. 1998. "By Carrot and by Rod: Extrinsic Motivation and English Language Attainment of Hong Kong Tertiary Students." In *Language in Hong Kong at Century's End*, edited by Martha Pennington, 285–301. Hong Kong: Hong Kong University Press.

Liu, Dilin. 2008. "Linking Adverbials: An Across-Register Corpus Study and Its Implications." *International Journal of Corpus Linguistics* 13 (4): 491–518. doi:10.1075/ijcl.13.4.05liu.

Ljung, Magnus. 1996. "The Use of Modals in British and American Newspapers." In *Words: Proceedings of an International Symposium, Lund, 25-26 August 1995*, edited by Jan Svartvik, 159–79. Stockholm: Almquist & Wiksell.

———. 1997. "Text Complexity in British and American Newspapers." In *Studies in English Language and Teaching*, edited by Jan Aarts, Inge d. Mönnink, and Herman Wekker, 75–83. Amsterdam: Rodopi.

———. 2000. "Newspaper Genres and Newspaper English." In *English Media Texts - Past and Present*, edited by Friedrich Ungerer, 131–49. Pragmatics & Beyond New Series 80. Amsterdam: John Benjamins.

Llewellyn, John, Greg Hancock, Michael Kirst, and Karl Roeloffs. 1982. *A Perspective on Education in Hong Kong: Report by a Visiting Panel.* Hong Kong: Government Printer.

Locke, Terry. 2004. *Critical Discourse Analysis.* London: Continuum.

Louwerse, Max, Philip M. McCarthy, Danielle S. McNamara, and Arthur C. Graesser. 2004. “Variation in Language and Cohesion across Written and Spoken Registers.” In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, edited by Ken Forbus, Dedre Gentner, and Terry Regier, 843–48. Mahwah, NJ: Lawrence Erlbaum.

Lovell, Julia. 2011. *The Opium War: Drugs, Dreams and the Making of China.* London: Picador.

Lowell, Lawrence. 1914. *Public Opinion and Popular Government.* New York: Longman, Green & Co.

Lowenberg, Peter H. 1992. “Testing English as a World Language: Issues in Assessing Non-Native Proficiency.” In *The Other Tongue: English across Cultures*, edited by Braj B. Kachru. 2nd ed., 108–21. Urbana and Chicago: University of Illinois Press.

Lüger, Heinz-Helmut. 1995. *Pressesprache.* 2nd ed. Germanistische Arbeitshefte 28. Tübingen: Niemeyer.

Luhmann, Niklas. 2000. *The Reality of the Mass Media.* Stanford: Stanford University Press.

Luke, Kang K. 1998. “Why Two Languages Might Be Better Than One: Motivations of Language Mixing in Hong Kong.” In *Language in Hong Kong at Century’s End*, edited by Martha Pennington, 145–59. Hong Kong: Hong Kong University Press.

Luke, Kang K., and Jack C. Richards. 1982. “English in Hong Kong: Functions and Status.” *English World-Wide* 3 (1): 47–64.

Lukin, Annabelle. 2010. “'News' and 'Register': A Preliminary Investigation.” In *Appliable Linguistics: Reclaiming the Place of Language in Linguistics*, edited by Ahmar Mahboob and Naomi Knight, 92–113. London: Continuum.

Lukin, Annabelle, Alison Moore, Maria Herke, Rebekah Wegener, and Canzhong Wu. 2011. “Halliday’s Model of Register Revisited and Explored.” *Linguistics and the Human Sciences* 4 (2): 187–213. doi:10.1558/lhs.v4i2.187.

Lupton, Deborah. 1993. “AIDS Risk and Heterosexuality in the Australian Press.” *Discourse & Society* 4 (3): 307–28. doi:10.1177/0957926593004003002.

Macalister, John. 2001. “Introducing a New Zealand Newspaper Corpus.” *New Zealand English Journal* 15: 35–41.

Machera, Mumbi. 2001. “Domestic Violence in Kenya: A Survey of Newspaper Reports.” In *Men, Women and Violence: a Collection of Papers from Codesria Gender Institute 1997*, edited by Felicia Oyekanmi, 27–28. Dakar: Council for the Development of Social Science Research in Africa (CODESRIA).

Macken-Horarik, Mary. 2003. “A Telling Symbiosis in the Discourse of Hatred: Multimodal News Texts about the "Children Overboard" Affair.” *Australian Review of Applied Linguistics* 26 (2): 1–16.

Maddalena, Antonello, and Isabel A. Belmonte. 2011. “Unveiling the Writer-Reader Interaction in Peninsular Spanish and American English Newspaper Editorials: A Cross Linguistic Study.” *Journal of Pragmatics* 43 (3): 891–903. doi:10.1016/j.pragma.2010.09.026.

Mahlberg, Michaela. 2009. “Local Textual Functions of Move in Newspaper Story Patterns.” In *Exploring the Lexis Grammar Interface*, edited by Ute Römer and Rainer Schulze, 265–87. Studies in Corpus Linguistics 35. Amsterdam: John Benjamins.

Mahlberg, Michaela, and Matthew B. O'Donnell. 2008. “A Fresh View of the Structure of Hard News Stories.” In *Online Proceedings of the 19th European Systemic Functional Linguistics Conference and Workshop, Saarbrücken, 23-25 July 2007*, edited by Stella Neumann and Erich Steiner. http://scidok.sulb.uni-saarland.de/volltexte/2008/1700. Accessed November 04, 2015.

Makkonen-Craig, Henna. 1999. “Speech Quotations in Newspapers as a Form of Language Use.” *SKY Journal of Linguistics* 12: 111–44.

Maley, Alan. 1985. “The Most Chameleon of Languages.” *English Today* 1 (1): 30–33. doi:10.1017/S0266078400013122.

Malinowski, Bronislaw. 1935. *Coral Gardens and Their Magic: A Study of the Methods of Tilling the Soil and of Agricultural Rites in the Trobriand Islands Vol 2: The Language of Magic and Gardening.* London: Allen & Unwin.

Manion, James. 1981. “History of Newspapers in North Queensland: The Clem Lack Memorial Oration Delivered on 24 March 1983.” *Royal Historical Society of Queensland* 11 (4): 139–51.

Manning, Peter. 2003. “Arabic and Muslim People in Sydney's Daily Newspapers, Before and After September 11.” *Media International Australia* 109 (1): 50–70.

Marchi, Anna. 2010. “‘The Moral in the Story’: A Diachronic Investigation of Lexicalised Morality in the UK Press.” *Corpora* 5 (2): 161–89. doi:10.3366/cor.2010.0104.

Marks, Robert. 2012. *China.* Lanham Md: Rowman & Littlefield.

Martin, James R. 1984. “Language, Register and Genre.” In *Children Writing: Reader*, edited by Frances Christie, 21–29. ECT418 Language Studies. Geelong, Victoria: Deakin University.

———. 1992. *English Text: System and Structure.* Philadelphia: John Benjamins.

———. 2004. “Positive Discourse Analysis: Solidarity and Change.” *Revista Canaria de Estudios Ingleses* 49: 179–200.

Martin, James R., and David Rose. 2008. *Genre Relations: Mapping Culture.* London: Equinox.

Martin, James R., and Peter R. R. White. 2005. *The Language of Evaluation: Appraisal in English.* Basingstoke: Palgrave Macmillan.

Masumoto, Tomoko. 2011. “The Semantic Dimensions of an International Story: the 'Ehime Maru' Incident.” *ETC: A Review of General Semantics* 68 (2): 205–13.

Mathesius, Vilém. 1929. “Zur Satzperspektive im modernen English.” *Archiv für das Studium der neueren Sprachen und Literaturen* 155: 202–10.

Matsuda, Aya, and Patricia Friedrich. 2011. “English as an International Language: A Curriculum Blueprint.” *World Englishes* 30 (3): 332–44. doi:10.1111/j.1467-971X.2011.01717.x.

Matsumoto, David, Hyisung C. Hwang, and Vincent A. Sandoval. 2015. “Cross-Language Applicability of Linguistic Features Associated with Veracity and Deception.” *Journal of Police and Criminal Psychology* 30 (4): 229–41. doi:10.1007/s11896-014-9155-0.

Matthews, Julian, and Andy R. Brown. 2012. “Negatively Shaping the Asylum Agenda? The Representational Strategy and Impact of a Tabloid News Campaign.” *Journalism* 13 (6): 802–17. doi:10.1177/1464884911431386.

Matu, Peter M., and Hendrik J. Lubbe. 2007. “Investigating Language and Ideology: A Presentation of the Ideological Square and Transitivity in the Editorials of Three Kenyan Newspapers.” *Journal of Language and Politics* 6 (3): 401–18. doi:10.1075/jlp.6.3.07mat.

Mazrui, Alamin, and Ali Mazrui. 1996. “A Tale of Two Englishes: The Imperial Language in Post-Colonial Kenya and Uganda.” In *Post-Imperial English: Status Change in Former British and American Colonies, 1940-1990*, edited by Joshua A. Fishman, Andrew W. Conrad, and Alma Rubal-Lopez, 271–302. New York: De Gruyter Mouton.

McArthur, Tom. 1987. “The English Languages?” *English Today* 3 (3): 9–13. doi:10.1017/S0266078400013511.

———. 1992. “Models of English.” *English Today* 8 (4): 12–21. doi:10.1017/S0266078400006684.

———. 1998. *The English Languages.* Cambridge: Cambridge University Press.

McEnery, Tony, Mark McGlashan, and Robbie Love. 2015. "Press and Social Media Reaction to Ideologically Inspired Murder: The Case of Lee Rigby." *Discourse & Communication* 9 (2): 237–59. doi:10.1177/1750481314568545.

McFarland, Katherine. 2011. "Media Influence and Frame Diversity in the Debate over Same-Sex Marriage." *The Communication Review* 14 (4): 255–78. doi:10.1080/10714421.2011.624002.

McKay, Jennifer M. 1983. "Newspaper Reporting of Bush-Fire Disaster in South-Eastern Australia-Ash Wednesday 1983." *Disasters* 7 (4): 283–90. doi:10.1111/j.1467-7717.1983.tb00836.x.

McLuhan, Marshall. (1964) 2001. *Understanding Media: The Extensions of Man.* London: Routledge.

McQuail, Denis. 2005. *McQuail's Mass Communication Theory.* 5th ed. London: Sage Publications.

McQuarrie, Edward F., and David G. Mick. 1996. "Figures of Rhetoric in Advertising Language." *Journal of Consumer Research* 22 (4): 424–38. doi:10.1086/209459.

Mead, George H. (1934) 1967. *Mind, Self and Society from the Standpoint of a Social Behaviorist.* Chicago: University of Chicago Press.

Meade, Aimee. 2013. "Bitcoins explained: A viable future currency, or an accident waiting to happen?" *Daily Mail*, April 5. Accessed November 10, 2016. http://www.dailymail.co.uk/money/news/article-2304451/Bitcoin-bubble-explained-A-viable-future-currency-investment-accident-waiting-happen.html.

Meadows, Michael. 2000. "Deals and Victories: Newspaper Coverage of Native Title in Australia and Canada." *Australian Journalism Review* 22 (1): 81–105.

Meeks, Lindsey. 2012. "Is She "Man Enough"? Women Candidates, Executive Political Offices, and News Coverage." *Journal of Communication* 62 (1): 175–93. doi:10.1111/j.1460-2466.2011.01621.x.

Mehler, Alexander, and Reinhard Köhler. 2007. "Introduction: Machine Learning in a Semiotic Perspective." In *Aspects of Automatic Text Analysis*, edited by Alexander Mehler and Reinhard Köhler, 1–29. Studies in Fuzziness and Soft Computing 209. Berlin: Springer.

Mei, Zhaoyang, Ren Zhang, and Baixiang Yu. 2015. "An Appraisal Analysis of Reports about Chinese Military Affairs in the 'New York Times'." In *Corpus Linguistics in Chinese Contexts*, edited by Bin Zou, Simon Smith, and Michael Hoey, 178–200. New Language Learning and Teaching Environments. Basingstoke: Palgrave Macmillan.

Meinl, Marja E. 2014. *Electronic Complaints: An Empirical Study on British English and German Complaints on EBay.* Kulturen - Kommunikation - Kontakte 18. Berlin: Frank & Timme.

Mersch, Dieter. 2006. *Medientheorien zur Einführung.* Hamburg: Junius.

Mesthrie, Rajend. 2006. "World Englishes and the Multilingual History of English." *World Englishes* 25 (3-4): 381–90. doi:10.1111/j.1467-971X.2006.00477.x.

Mesthrie, Rajend, and Rakesh M. Bhatt. 2008. *World Englishes: The Study of New Linguistic Varieties.* Cambridge: Cambridge University Press.

Michieka, Martha M. 2005. "English in Kenya: A Sociolinguistic Profile." *World Englishes* 24 (2): 173–86. doi:10.1111/j.1467-971X.2005.00402.x.

———. 2009. "Expanding Circles within the Outer Circle: The Rural Kisii in Kenya." *World Englishes* 28 (3): 352–64. doi:10.1111/j.1467-971X.2009.01597.x.

Mieroop, Dorien van de. 2005. “An Integrated Approach of Quantitative and Qualitative Analysis in the Study of Identity in Speeches.” *Discourse & Society* 16 (1): 107–30. doi:10.1177/0957926505048232.

Ming, Liu. 2015. “Scapegoat or Manipulated Victim? Metaphorical Representations of the Sino-US Currency Dispute in Chinese and American Financial News.” *Text & Talk* 35 (3). doi:10.1515/text-2015-0003.

Moag, Rodney. 1983. “The Life Cycle of Non-Native Englishes: A Case Study.” In *The Other Tongue: English across Cultures*, edited by Braj B. Kachru. 1st ed., 270–88. Oxford: Pergamon Press.

Modiano, Marko. 1999. “International English in the Global Village.” *English Today* 15 (2): 22–28. doi:10.1017/S026607840001083X.

Moe, Marija Z. 2010. “Register Shifts in Translations of Popular Fiction from English into Slovene.” In *Why Translation Studies Matters*, edited by Daniel Gile, Gyde Hansen, and Nike K. Pokorn, 125–36. Benjamins Translation Library 88. Amsterdam: John Benjamins.

Moise, Edwin E. 1986. *Modern China: A History.* London: Longman.

Mollin, Sandra. 2007. “New Variety or Learner English? Criteria for Variety Status and the Case of Euro-English.” *English World-Wide* 28 (2): 167–85.

Moore, Stephen, and Suksiri Bounchan. 2010. “English in Cambodia: Changes and Challenges.” *World Englishes* 29 (1): 114–26. doi:10.1111/j.1467-971X.2009.01628.x.

Mouton, Andre. 2013. “Will consumers embrace wearable tech?” *USA Today*, July 15. Accessed November 10, 2016. http://www.usatoday.com/story/tech/2013/07/15/wear-tech-google-glass-iwatch/2517379/.

Mufwene, Salikoko. 1994. “New Englishes and Criteria for Naming Them.” *World Englishes* 13 (1): 21–31. doi:10.1111/j.1467-971X.1994.tb00280.x.

———. 2001. *The Ecology of Language Evolution.* Cambridge: Cambridge University Press.

———. 2015. “Colonization, Indigenization, and the Differential Evolution of English: Some Ecological Perspectives.” *World Englishes* 34 (1): 6–21. doi:10.1111/weng.12129.

Mukherjee, Joybrato, and Stefan T. Gries. 2009. “Collostructional Nativisation in New Englishes: Verb-Construction Associations in the International Corpus of English.” *English World-Wide* 30 (1): 27–51. doi:10.1075/eww.30.1.03muk.

Mukherjee, Joybrato, and Sebastian Hoffmann. 2006. “Describing Verb-Complementational Profiles of New Englishes: A Pilot Study of Indian English.” *English World-Wide* 27 (2): 147–73. doi:10.1075/eww.27.2.03muk.

Mukherjee, Joybrato, and Marco Schilk. 2008. “Verb-Complementational Profiles across Varieties of English: Comparing Verb Classes in Indian English and British English.” In *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*, edited by Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta, and Minna Korhonen, 163–81. Amsterdam: John Benjamins.

Murata, Kumiko. 2007. “Pro- and Anti-Whaling Discourses in British and Japanese Newspaper Reports in Comparison: A Cross-Cultural Perspective.” *Discourse & Society* 18 (6): 741–64. doi:10.1177/0957926507082194.

Murray, Craig, Katy Parry, Piers Robinson, and Peter Goddard. 2008. “Reporting Dissent in Wartime: British Press, the Anti-War Movement and the 2003 Iraq War.” *European Journal of Communication* 23 (1): 7–27. doi:10.1177/0267323107085836.

Mwangi, Serah. 2003. *Prepositions in Kenyan English: A Corpus-Based Study in Lexico-Grammatical Variation.* Sprache & Kultur. Aachen, Chemnitz: Shaker.

Nabea, Wendo. 2009. “Language Policy in Kenya: Negotiation with Hegemony.” *The Journal of Pan African Studies* 3 (1): 121–38.

Nelson, Gerald. 2006. “The Core and Periphery of World Englishes: A Corpus-Based Exploration.” *World Englishes* 25 (1): 115–29.

Nesselhauf, Nadja. 2009. “Co-Selection Phenomena across New Englishes: Parallels (and Differences) to Foreign Learner Varieties.” *English World-Wide* 30 (1): 1–26.

Nessheim, Ragnhild M. V. 2007. “British National Newspapers and the Blair Government. Part I: The Easy Ride, 1997 – 2001.” *English Studies* 88 (6): 653–69. doi:10.1080/00138380701706468.

———. 2008. “British National Newspapers and the Blair Government 1997 – 2007. Part II: How Tony Blair Became “Tony B. Liar” and Nearly Caused Another Realignment of the Murdoch Press.” *English Studies* 89 (1): 12–38. doi:10.1080/00138380701771025.

Neumann, Stella. 2012. “Applying Register Analysis to Varieties of English.” In *Anglistentag 2011 Freiburg: Proceedings*, edited by Monika Fludernik and Benjamin Kohlmann, 75–94. Trier: WVT.

———. 2013. *Contrastive Register Variation: A Quantitative Approach to the Comparison of English and German.* Berlin: De Gruyter Mouton.

Neumann, Stella, and Jennifer Fest. 2016. “Cohesive Devices across Registers and Varieties: The Role of Medium in English.” In *Variational Text Linguistics: Revisiting Register in English*, edited by Christoph Schubert and Christina Sanchez-Stockhammer, 195–220. Topics in English Linguistics 90. Berlin: De Gruyter Mouton.

Nevala, Minna. 2004. “Accessing Politeness Aaxes: Forms of Address and Terms of Reference in Early English Correspondence.” *Journal of Pragmatics* 36 (12): 2125–60. doi:10.1016/j.pragma.2004.02.001.

Nielsen Company. 2015. “Nielsen Estimates 116.4 Million TV Homes in the U.S. for the 2015-16 TV Sseason.” Accessed November 19, 2015. http://www.nielsen.com/us/en/insights/news/2015/nielsen-estimates-116-4-million-tv-homes-in-the-us-for-the-2015-16-tv-season.html.

Noël, Dirk, and Johan van der Auwera. 2015. “Recent Quantitative Changes in the Use of Modals and Qquasi-Modals in the Hong Kong, British and American Printed Press.” In *Grammatical Change in English World-Wide*. Vol. 67, edited by Peter Collins, 437–64. Studies in Corpus Linguistics 67. Amsterdam: John Benjamins.

Noelle-Neumann, Elisabeth. 1995. “Wirkung der Massenmedien auf die Meinungsbildung.” In *Fischer Lexikon Publizistik. Massenkommunikation*, edited by Elisabeth Noelle-Neumann, Winfried Schulz, and Jürgen Wilke, 518–71. Frankfurt am Main: Fischer.

Nolan, David, Karen Farquharson, Violeta Politoff, and Timothy Marjoribanks. 2011. “Mediated Multiculturalism: Newspaper Representations of Sudanese Migrants in Australia.” *Journal of Intercultural Studies* 32 (6): 655–71. doi:10.1080/07256868.2011.618109.

Nunan, David. 2007. *What Is This Thing Called Language?* London: Palgrave Macmillan.

———. 2008. “Exploring Genre and Register in Contemporary English.” *English Today* 24 (2): 56–61.

Nyabuga, George, and Nancy Booker. 2013. “Mapping Digital Media: Kenya: A Report by the Open Society Foundation.” Accessed November 08, 2015. https://www.opensocietyfoundations.org/sites/default/files/mapping-digital-media-kenya-20130321.pdf.

O'Donnell, Shannon. 2013. "The truth about travelling alone as a woman." *The Courier Mail*, September 29. Accessed November 10, 2016. http://www.courier-mail.com.au/travel/travel-advice/the-truth-about-travelling-alone-as-a-woman/story-fnihr5wo-1226729177996#&mm-breached.

Ochilo, Polycarp J. O. 1993. "Press Freedom and the Role of the Media in Kenya." *Africa Media Review* 7 (3): 19–33.

Ogola, George. 2011. "The Political Economy of the Media in Kenya: from Kenyatta's Nation-Building Press to Kibaki's Local-Language FM Radio." *Africa Today* 57 (3): 77–95.

Ongong'a, Stephen O. 2008. "The Shaping of News Values of Young Journalists in Kenya." *African Communication Research* 1 (2): 159–84.

Oostdijk, Nelleke. 1988. "A Corpus Linguistic Approach to Linguistic Variation." *Literary and Linguistic Computing* 3 (1): 12–25.

Orr, Aleisha. 2012. "Teens charged over armed robbery." *The Sydney Morning Herald*, July 12. Accessed November 10, 2016. http://www.smh.com.au/wa-news/teens-charged-over-armed-robbery-20120712-21xyp.html.

Oxford English Dictionary. 2016. "OED Online." Accessed January 17, 2016. http://www.oed.com/.

Ozturk, Ismet. 2007. "The Textual Organisation of Research Article Introductions in Applied Linguistics: Variability within a Single Discipline." *English for Specific Purposes* 26 (1): 25–38. doi:10.1016/j.esp.2005.12.003.

Pan, Xiaping. 2002. "Consensus behind Disputes: A Critical Discourse Analysis of the Media Coverage of the Right-of-Abode Issue in Postcolonial Hong Kong." *Media, Culture & Society* 24 (1): 49–68. doi:10.1177/016344370202400103.

Pape, Susan, and Sue Featherstone. 2005. *Newspaper Journalism: A Practical Introduction.* London: Sage Publications.

Papper, Robert A. 2013. *Broadcast News and Writing Stylebook.* 5th ed. Boston: Pearson.

Partington, Alan. 2010. "Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS) on UK Newspapers: An Overview of the Project." *Corpora* 5 (2): 83–108. doi:10.3366/cor.2010.0101.

Peacock, Matthew. 2015. "Stance Adverbials in Research Writing." *Ibérica* 29: 35–61.

Peelo, Moira, Brian Francis, Keith Soothill, Jayn Pearson, and Elizabeth Ackerley. 2004. "Newspaper Reporting and the Public Construction of Homicide." *British Journal of Criminology* 44 (2): 256–75. doi:10.1093/bjc/44.2.256.

Pellechia, Marianne G. 1997. "Trends in Science Coverage: A Content Analysis of Three US Newspapers." *Public Understanding of Science* 6 (1): 49–68. doi:10.1088/0963-6625/6/1/004.

Pennington, Martha. 1995. "Language Diversity in Bilingualism: Preliminary Speculations on Varieties of Hong Kong English." *Language in Education Journal* 1 (1): 1–19.

———. 1998a. "Colonial's Aftermath in Asia: A Snapshot View of Bilingualism in Hong Kong." *Hong Kong Journal of Applied Linguistics* 3 (1): 1–6.

———. 1998b. "Introduction: Perspectives on Language in Hong Kong at Century's End." In *Language in Hong Kong at Century's End*, edited by Martha Pennington, 3–40. Hong Kong: Hong Kong University Press.

Pennington, Martha, and Francis Yue. 1994. "English and Chinese in Hong Kong: Pre-1997 Language Attitudes." *World Englishes* 13 (1): 1–20.

Pennycook, Alastair. 1994. “The Politics of Pronouns.” *English Language Teaching Journal* 48 (2): 173–78. doi:10.1093/elt/48.2.173.

———. 1998. *English and the Discourses of Colonialism.* The Politics of Language. London: Routledge.

Pérez-Sobrino, Paula. 2013. “Personification and Ideology in the American Media Coverage of the Iranian Green Revolution.” *Text & Talk* 33 (2): 233–58. doi:10.1515/text-2013-0011.

Peterlin, Agnes P. 2005. “Text-Organising Metatext in Research Articles: An English–Slovene Contrastive Analysis.” *English for Specific Purposes* 24 (3): 307–19. doi:10.1016/j.esp.2004.11.001.

Peters, Pam, Peter Collins, David Blair, and Alison Brierley. 1988. “The Australian Corpus Project: Findings on the Distribution of Some Functional Variants in the Australian Press.” *Australian Review of Applied Linguistics* 11 (1): 22–33.

Pickering, Sharon. 2001. “Common Sense and Original Deviancy: News Discourses and Asylum Seekers in Australia.” *Journal of Refugee Studies* 14 (2): 169–86. doi:10.1093/jrs/14.2.169.

Pierson, Herbert D., and Michael H. Bond. 1982. “How Do Chinese Bilinguals Respond to Variations of Interviewer Language and Ethnicity?” *Journal of Language and Social Psychology* 1 (2): 123–39. doi:10.1177/0261927X8200100203.

Pirkis, Jane E., Philip M. Burgess, Catherine Francis, R. W. Blood, and Damien J. Jolley. 2006. “The Relationship between Media Reporting of Suicide and Actual Suicide in Australia.” *Social Science & Medicine* 62 (11): 2874–86. doi:10.1016/j.socscimed.2005.11.033.

Platt, John. 1986. “English in Singapore, Malaysia, and Hong Kong.” In *English as a World Language*, edited by Richard W. Bailey and Manfred Görlach. 4th ed., 384–414. Ann Arbor: The University of Michigan Press.

Platt, John, Heidi Weber, and Mian L. Ho. 1984. *The New Englishes.* London, Boston, Melbourne, Henley: Routledge and Kegan Paul.

Plesner, Ursula. 2012. “The Performativity of "Media Logic" in the Mass Mediation of Science.” *Public Understanding of Science* 21 (6): 674–88. doi:10.1177/0963662510385309.

Pollak, Senja, Roel Coesemans, Walter Daelemans, and Nada Lavrač. 2011. “Detecting Contrast Patterns in Newspaper Articles by Combining Discourse Analysis and Text Mining.” *Pragmatics* 21 (4): 647–83. doi:10.1075/prag.21.4.07pol.

Poon, Anita Y. K. 2009. “A Review of Research in English Language Education in Hong Kong in the Past 25 Years: Reflections and the Way Forward.” *Educational Research Journal* 24 (1): 7–40.

Potts, Amanda, Monika Bednarek, and Helen Caple. 2015. “How Can Computer-Based Methods Help Researchers to Investigate News Values in Large Datasets? A Corpus Linguistic Study of the Construction of Newsworthiness in the Reporting on Hurricane Katrina.” *Discourse & Communication* 9 (2): 149–72. doi:10.1177/1750481314568548.

Powell, James. 2001. “The Allure of Foreign Affairs.” *The Guardian*, October 30. Accessed April 17, 2016. http://www.theguardian.com/media/2001/oct/30/marketingandpr.

Poynton, Cate. 1989. *Language and Gender: Making the Difference.* Oxford: Oxford University Press.

Precht, Kristen. 1998. “A Cross-Cultural Comparison of Letters of Recommendation.” *English for Specific Purposes* 17 (3): 241–65. doi:10.1016/S0889-4906(97)00012-4.

Pullman, Laura. 2013. “Aaargh! The world's silliest haircut is back and it's huge on the catwalk. But can a real woman pull off a MULLET?” *Daily Mail*, March 27. Accessed November 10,

2016. http://www.dailymail.co.uk/femail/article-2300232/Can-real-woman-pull-mullet-haircut.html.

Putseys, Yvan. 1993. “The World's Languages – How Many Sspeakers?” *English Today* 9 (4): 4–12. doi:10.1017/S0266078400007239.

Pyykkö, Riitta. 2002. “Who Is ‘Us’ in Russian Political Discourse.” In *Us and Others: Social Identities across Languages, Discourses and Cultures*, edited by Anna Duszak, 233–48. Pragmatics & Beyond New Series 98. Amsterdam: John Benjamins.

Quinn, Stephen, and Vincent F. Filak, eds. 2005. *Convergent Journalism - An Introduction: Writing and Producing across Media.* Oxford: Focal Press.

Quirk, Randolph. (1962) 1968. *The Use of English.* 2nd ed. London: Longman.

———. 1981. “International Communication and the Concept of Nuclear English.” In *English for Cross-Cultural Communication*, edited by Larry E. Smith, 151–65. London: Palgrave Macmillan.

———. 1985. “The English Language in a Global Context.” In *English in the World: Teaching and Learning the Language and Literatures; Papers of an International Conference Entitled 'Progress in English Studies' Held in London, 17 - 21 September 1984 to Celebrate the Fiftieth Anniversary of the British Council and Its Contribution to the Field of English Studies over Fifty Years*, edited by Randolph Quirk and Henry G. Widdowson, 1–6. Cambridge: Cambridge University Press.

———. 1990. “Language Varieties and Standard Language.” *English Today* 6 (1): 3–10. doi:10.1017/S0266078400004454.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language.* 2nd ed. Harlow: Longman.

Rafail, Patrick, Edward T. Walker, and John D. McCarthy. 2015. “Protests on the Front Page: Media Salience, Institutional Dynamics, and Coverage of Collective Action in the New York Times, 1960-1995.” *Communication Research.* doi:10.1177/0093650215608239.

Raguseo, Carla. 2010. “Twitter Fiction: Social Networking and Microfiction in 140 Characters.” *Teaching English as a Second or Foreign Language* 13 (4). http://tesl-ej.org/pdf/ej52/int.pdf. Accessed May 12, 2015.

Ramson, William S. 1989. “Regionalism in Australian English: The Value of Newspaper Evidence.” *Australian Journal of Linguistics* 9 (1): 73–83. doi:10.1080/07268608908599412.

Randall, David. 2000. *The Universal Journalist.* 2nd ed. London: Pluto Press.

Reah, Danuta. 2003. *The Language of Newspapers.* 2nd ed. London: Routledge.

Reaser, Jeffrey. 2003. “A Quantitative Approach to (Sub)registers: The Case of "Sports Announcer Talk".” *Discourse Studies* 5 (3): 303–21. doi:10.1177/14614456030053002.

Reinemann, Carsten, James Stanyer, Sebastian Scherr, and Guido Legnante. 2012. “Hard and Soft News: A Review of Concepts, Operationalizations and Key Findings.” *Journalism* 13 (2): 221–39. doi:10.1177/1464884911427803.

Renkema, Jan. 2004. *Introduction to Discourse Studies.* Amsterdam: John Benjamins.

Reporters without Borders. 2015. “World Press Freedom Index.” Accessed February 04, 2016. http://index.rsf.org/#!/.

Reuters. 2000. “Reuters Corpus.” Accessed February 28, 2016. http://about.reuters.com/researchandstandards/corpus/.

Reynolds, Mike, and Giovanna Cascio. 1999. “The Use of Contracted Forms in British Newspapers.” In *English via Various Media*. Vol. 62, edited by Hans-Jürgen Diller, Erwin Otto, and

Gert Stratmann, 179–200. Anglistik und Englischunterricht 62. Heidelberg: Universitätsverlag C. Winter.

Richards, Jack C., and Mary Tay. 1981. "Norm and Variability in Language Use and Language Learning." In *English for Cross-Cultural Communication*, edited by Larry E. Smith, 40–56. London: Palgrave Macmillan.

Richards, Stephen. 1998. "Learning English in Hong Kong: Making Connections between Motivation, Language Use, and Strategy Choice." In *Language in Hong Kong at Century's End*, edited by Martha Pennington, 303–28. Hong Kong: Hong Kong University Press.

Richardson, John E. 2004. *(Mis)representing Islam: The Racism and Rhetoric of British Broadsheet Newspapers.* Discourse Approaches to Politics, Society, and Culture 9. Amsterdam: John Benjamins.

———. 2009. *Analysing Newspapers: An Approach from Critical Discourse Analysis.* 3rd ed. Basingstoke: Palgrave Macmillan.

Ricketson, Matthew. 2004. *Writing Feature Stories: How to Research and Write Newspaper and Magazine Articles.* Crows Nest, NSW: Allen & Unwin.

Robertson, James. 2013. "Four-car crash leaves three seriously injured." *The Sydney Morning Herald*, April 8. Accessed November 10, 2016. http://www.smh.com.au/nsw/fourcar-crash-leaves-three-seriously-injured-20130408-2hh6h.html.

Rodríguez, María J. G. 2006. "Tracing Context in the Discourse of the Media: Features of Language-in-Use in the British Press." *Revista Alicantina de Estudios Ingleses* 19: 149–68.

Rojas, John-Paul F. 2012. "Ukip takes second place in by-elections." *The Daily Telegraph*, November 30. Accessed November 09, 2016. http://www.telegraph.co.uk/news/politics/9713243/Ukip-takes-second-place-in-by-elections.html.

Rojas-Lizana, Isolda. 2011. "Justifying and Condemning Sexual Discrimination in Everyday Discourse: Letters to the Editor in the Australian Local Press." *Journal of Pragmatics* 43 (2): 663–76. doi:10.1016/j.pragma.2010.09.007.

Romaine, Suzanne. 1992. "English: from Village to Global Village." In *English in Its Social Contexts: Essays in Historical Sociolinguistics*, edited by Tim Machan and Charles Scott, 253–60. Oxford: Open University Press.

Rooy, Bertus van, Lize Terblanche, Christoph Haase, and Josef J. Schmied. 2010. "Register Differentiation in East African English: A Multidimensional Study." *English World-Wide* 31 (3): 311–49. doi:10.1075/eww.31.3.04van.

Ruiying, Yang, and Desmond Allison. 2003. "Research Articles in Applied Linguistics: Moving from Results to Conclusions." *English for Specific Purposes* 22 (4): 365–85. doi:10.1016/S0889-4906(02)00026-1.

———. 2004. "Research Articles in Applied Linguistics: Structures from a Functional Perspective." *English for Specific Purposes* 23 (3): 264–79. doi:10.1016/S0889-4906(03)00005-X.

Rushdie, Salman. 1982. "The Empire Writes Back with a Vengeance." *The Times*, July 3.

Ryan, J. S. 2005. "Word, Truth and Gender Bending in Australia: The Drive to Amuse in Story or Report, Rather Than to Inform." *Australian Folklore: A Yearly Journal of Folklore Studies* 20: 99–102.

Ryan, Michael. 2004. "Framing the War against Terrorism: US Newspaper Editorials and Military Action in Afghanistan." *International Communication Gazette* 66 (5): 363–82. doi:10.1177/0016549204045918.

Rydén, Mats. 2008. "Noun-Name Collocations in British English Newspaper Language." *Studia Neophilologica* 47 (1): 14–39. doi:10.1080/00393277508587609.

Salager-Meyer, Françoise. 1999. "Referential Behavior in Scientific Writing: A Diachronic Study (1810–1995)." *English for Specific Purposes* 18 (3): 279–305. doi:10.1016/S0889-4906(97)00042-2.

Samraj, Betty. 2002. "Introductions in Research Articles: Variations across Disciplines." *English for Specific Purposes* 21 (1): 1–17. doi:10.1016/S0889-4906(00)00023-5.

Sand, Andrea. 2004. "Shared Morpho-Syntactic Features of Contact Varieties: Article Use." *World Englishes* 23 (2): 281–98.

Santana, Arthur D. 2014. "Virtuous or Vitriolic: The Effect of Anonymity on Civility in Online Newspaper Reader Comment Boards." *Journalism Practice* 8 (1): 18–33. doi:10.1080/17512786.2013.813194.

Santorini, Beatrice. 1991. "Part-of-Speech Tagging Guidelines for the Penn Treebank Project." Accessed March 10, 2016. http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/Penn-Treebank-Tagset.pdf.

Schaefer, Todd. 2004. "Framing the US Embassy Bombings and September 11 Attacks in African and US Newspapers." In *Framing Terrorism: The News Media, the Government and the Public*, edited by Pippa Norris, Montague Kern, and Marion Just, 93–112. New York: Routledge.

Schäfers, Bernhard. 2002. "Soziales Handeln und seine Grundlagen: Normen, Werte, Sinn." In *Einführung in die Hauptbegriffe der Soziologie*, edited by Hermann Korte and Bernhard Schäfers. 6th ed., 25–43. Einführungskurs Soziologie 1. Opladen: Leske + Budrich.

Schenk, Michael. 2007. *Medienwirkungsforschung.* 3rd ed. Tübingen: Mohr Siebeck.

Scherr, Albert. 2002. "Sozialisation, Person, Individuum." In *Einführung in die Hauptbegriffe der Soziologie*, edited by Hermann Korte and Bernhard Schäfers. 6th ed., 45–66. Einführungskurs Soziologie 1. Opladen: Leske + Budrich.

Schmid, Helmut. 1994. "Probabilistic Part-of-Speech Tagging Using Decision Trees." Proceedings of International Conference on New Methods in Language Processing, Manchester, UK, 1994.

———. 1995. "Improvements in Part-of-Speech Tagging with an Application to German." Proceedings of the ACL SIGDAT-Workshop, Dublin, Ireland, 1995.

Schmied, Josef J. 1989a. "English in East Africa: Theoretical, Methodological and Practical Issues." In *English in East and Central Africa I*, edited by Josef J. Schmied, 7–37. Bayreuth African Studies Series 15. Altendorf: Gräbner.

———. 1989b. "Second-Language Varieties across the Indian Ocean." In *English in East and Central Africa I*, edited by Josef J. Schmied, 85–97. Bayreuth African Studies Series 15. Altendorf: Gräbner.

———. 1991a. *English in Africa: An Introduction.* London: Longman.

———. 1991b. "National and Subnational Features in Kenyan English." In *English around the World: Sociolinguistic Perspectives*, edited by Jenny Cheshire, 420–32. Cambridge: Cambridge University Press.

———. 2006. "East African Englishes." In *The Handbook of World Englishes*, edited by Braj B. Kachru, Yamuna Kachru, and Cecil L. Nelson, 188–202. Blackwell Handbooks in Linguistics. Malden, Mass: Blackwell.

Schmied, Josef J., and Diana Hudson-Ettle. 1996. "Analyzing the Style of East African Newspapers in English." *World Englishes* 15 (1): 103–13. doi:10.1111/j.1467971X.1996.tb00095.x.

Schneider, Edgar W. 2003. "The Dynamics of New Englishes: from Identity Construction to Dialect Birth." *Language* 79 (2): 233–81. doi:10.1353/lan.2003.0136.

———. 2007. *Postcolonial English: Varieties around the World.* Cambridge: Cambridge University Press.

———. 2014. "New Reflections on the Evolutionary Dynamics of World Englishes." *World Englishes* 33 (1): 9–32. doi:10.1111/weng.12069.

Schneider, Kristina. 1999. "Exploring the Roots of Popular English News Writing: A Preliminary Report on a Corpus-Based Project." In *English via Various Media*, edited by Hans-Jürgen Diller, Erwin Otto, and Gert Stratmann, 201–22. Anglistik und Englischunterricht 62. Heidelberg: Universitätsverlag C. Winter.

———. 2000. "The Emergence and Development of Headlines in English Newspapers." In *English Media Texts - Past and Present*, edited by Friedrich Ungerer, 45–65. Pragmatics & Beyond New Series 80. Amsterdam: John Benjamins.

———. 2012. "Rostock Newspaper Corpus." Accessed February 28, 2016. http://www.iaa.uni-rostock.de/lehrstuehle-abteilungen/sprachwissenschaft/rostock-newspaper-corpus/.

Schubert, Christoph. 2008. *Englische Textlinguistik: Eine Einführung.* Grundlagen der Anglistik und Amerikanistik 30. Berlin: Erich Schmidt Verlag.

———. 2016. "Introduction: Current Trends in Register Research." In *Variational Text Linguistics: Revisiting Register in English*, edited by Christoph Schubert and Christina Sanchez-Stockhammer, 1–16. Topics in English Linguistics 90. Berlin: De Gruyter Mouton.

Schultz, Tanjev. 1999. "Interactive Options in Online Journalism: A Content Analysis of 100 U.S. Newspapers." *Journal of Computer-Mediated Communication* 5 (1). doi:10.1111/j.1083-6101.1999.tb00331.x.

Schutz, Natassia. 2013. "How Specific Is English for Academic Purposes? A Look at Verbs in Business, Linguistics and Medical Research Articles." In *English Corpus Linguistics: Variation in Time, Space and Genre*, edited by Gisle Andersen and Kristin Bech, 237–57. Language and Computers: Studies in Practical Linguistics 77. Amsterdam: Rodopi.

Sciutto, James E. 1996. "China's Muffling of the Hong Kong Media." *The Annals of the American Academy of Political and Social Science* 547 (1): 131–43. doi:10.1177/0002716296547001011.

Scollon, Ron. 1997. "Attribution and Power in Hong Kong News Discourse." *World Englishes* 16 (3): 383–93. doi:10.1111/1467-971X.00072.

———. 2000. "Generic Variability in News Stories in Chinese and English: A Contrastive Discourse Study of Five Days' Newspapers." *Journal of Pragmatics* 32 (6): 761–91. doi:10.1016/S0378-2166(99)00092-2.

Scollon, Ron, and Suzie Scollon. 1997. "Point of View and Citation: Fourteen Chinese and English Versions of the 'Same' News Story." *Text - Interdisciplinary Journal for the Study of Discourse* 17 (1): 83–125. doi:10.1515/text.1.1997.17.1.83.

Scott, Claire E. 2011. "Exploring Diachronic Register Change in Reports of Armistice 1902–2003." *Linguistics and the Human Sciences* 4 (3): 241–64. doi:10.1558/lhs.v4i3.241.

Scott, Mike. 2001. "Mapping Key Words to Problem and Solution." In *Patterns of Text: In Honour of Michael Hoey*, edited by Mike Scott and Geoff Thompson, 109–27. Amsterdam: John Benjamins.

Seargeant, Philip. 2005. "Globalisation and Reconfigured English in Japan." *World Englishes* 24 (3): 309–19. doi:10.1111/j.0083-2919.2005.00412.x.

———. 2010. "Naming and Defining in World Englishes." *World Englishes* 29 (1): 97–113. doi:10.1111/j.1467-971X.2009.01627.x.

———. 2012. *Exploring World Englishes: Language in a Global Context.* London: Routledge.

Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language.* Cambridge: Cambridge University Press.

Semino, Elena. 2002. "A Sturdy Baby or a Derailing Train? Metaphorical Representations of the Euro in British and Italian Newspapers." *Text - Interdisciplinary Journal for the Study of Discourse* 22 (1): 107–39. doi:10.1515/text.2002.001.

Sewell, Andrew, and Jason Chan. 2010. "Patterns of Variation in the Consonantal Phonology of Hong Kong English." *English World-Wide* 31 (2): 138–61.

Seymour, Michael. 2004. "Ancient Mesopotamia and Modern Iraq in the British Press, 1980–20031." *Current Anthropology* 45 (3): 351–68. doi:10.1086/383004.

Sharma, Devyani. 2009. "Typological Diversity in New Englishes." *English World-Wide* 30 (2): 170–95. doi:10.1075/eww.30.2.04sha.

Shaw, Willard D. 1981. "Asian Student Attitudes towards English." In *English for Cross-Cultural Communication*, edited by Larry E. Smith, 108–22. London: Palgrave Macmillan.

Sheriff, Abdulrahman. 2013. "Coast side vows to boycott matches over officiating." *Daily Nation,* October 22. Accessed November 10, 2016. http://www.nation.co.ke/sports/football/Coast-side-vows-to-boycott-matches/-/1102/2042772/-/ygee1xz/-/index.html.

Shoemaker, Pamela J., and Akiba A. Cohen. 2006. *News around the World: Content, Practitioners, and the Public.* London: Routledge.

Sin, King-sui, and Derek Roebuck. 1996. "Language Engineering for Legal Transplantation: Conceptual Problems in Creating Common Law Chinese." *Language & Communication* 16 (3): 235–54. doi:10.1016/0271-5309(96)00017-1.

Singer, Jane B. 2006. "Stepping Back from the Gate: Online Newspaper Editors and the Co-Production of Content in Campaign 2004." *Journalism & Mass Communication Quarterly* 83 (2): 265–80. doi:10.1177/107769900608300203.

Skandera, Paul. 1999. "What Do We Really Know about Kenyan English? A Pilot Study in Research Methodology." *English World-Wide* 20 (2): 217–36. doi:10.1075/eww.20.2.02ska.

———. 2003. *Drawing a Map of Africa.* Tübingen: Narr.

Smirnova, Alla. 2012. "Argumentative Use of Reported Speech in British Newspaper Discourse." *Text & Talk* 32 (2): 235–53. doi:10.1515/text-2012-0012.

Smith, Ben J., Claire Ferguson, Jeanie McKenzie, Adrian Bauman, and Philip Vita. 2002. "Impacts from Repeated Mass Media Campaigns to Promote Sun Protection in Australia." *Health Promotion International* 17 (1): 51–60. doi:10.1093/heapro/17.1.51.

Smith, Emily. 2008. "Obama faces off against McCain." *The Sun*, September 27.

Smith, Larry E. 1992. "Spread of English and Issues of Intelligibility." In *The Other Tongue: English across Cultures*, edited by Braj B. Kachru. 2nd ed., 75–90. Urbana and Chicago: University of Illinois Press.

Smotrova, Tetyana. 2008. "The Concept of "Strange" in American and Ukrainian Newspaper Discourse." *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis* 13 (2): 197–221.

So, Daniel. 1992. "Language-Based Bifurcation of Secondary Education in Hong Kong: Past, Present and Future." In *Into the Twenty First Century: Issues of Language and Education in Hong Kong*, edited by Kang K. Luke, 69–95. Hong Kong: Linguistic Society of Hong Kong.

South China Morning Post. 1986. "Language Project Deserves Applause." November 1.

South China Morning Post. 1989. "'Worst' English Tag Should Be Eliminated. Editorial." February 27.

Springer, Ken. 2010. *Educational Research: A Contextual Approach.* Hoboken, N.J. John Wiley.

Stamou, Anastasia. 2001. “The Representation of Non-Protesters in a Student and Teacher Protest: A Critical Discourse Analysis of News Reporting in a Greek Newspaper.” *Discourse & Society* 12 (5): 653–80. doi:10.1177/0957926501012005005.

Starck, Nigel. 2004. “Writes of Passage: A Comparative Study of Newspaper Obituary Practice in Australia, Britain and the United States.” PhD Thesis, Flinders University Adelaide.

Steiner, Erich. 1991. *A Functional Perspective on Language, Action, and Interpretation. An Initial Approach with a View to Computational Modeling.* Berlin, New York: De Gruyter Mouton.

———. 2004. *Translated Texts: Properties, Variants, Evaluations.* Frankfurt am Main: Peter Lang.

Stirling, Lesley. 1987. “Language and Gender in Australian Newspapers.” In *Women and Language in Australian and New Zealand Society*, edited by Anne Pauwels, 108–28. Sydney: Australian Professional Publications.

Stotesbury, Hilkka. 2003. “Evaluation in Research Article Abstracts in the Narrative and Hard Sciences.” *Journal of English for Academic Purposes* 2 (4): 327–41. doi:10.1016/S1475-1585(03)00049-3.

Strang, Barbara. 1970. *A History of English.* London: Methuen.

Strevens, Peter. 1980. *Teaching English as an International Language: from Practice to Principle.* Oxford: Pergamon Press.

———. 1981. “Forms of English: An Analysis of the Variables.” In *English for Cross-Cultural Communication*, edited by Larry E. Smith, 1–14. London: Palgrave Macmillan.

———. 1992. “English as an International Language: Directions in the 1990s.” In *The Other Tongue: English across Cultures*, edited by Braj B. Kachru. 2nd ed., 27–47. Urbana and Chicago: University of Illinois Press.

Stubbs, Michael. 1983. *Discourse Analysis.* Oxford: Blackwell.

———. 1996. *Text and Corpus Analysis: Computer Assisted Studies of Language and Institutions.* Language in Society 23. Oxford: Blackwell.

Studer, Patrick. 2003. “Textual Structures in Eighteenth-Century Newspapers: A Corpus-Based Study of Headlines.” *Journal of Historical Pragmatics* 4 (1): 19–44. doi:10.1075/jhp.4.1.03stu.

Sure, Kembo. 1989. “Attitudes towards English among Kenyan Students.” In *English in East and Central Africa I*, edited by Josef J. Schmied, 39–61. Bayreuth African Studies Series 15. Altendorf: Gräbner.

Swales, John M. 1985. “ESP - the Heart of the Matter or the End of the Affair?” In *English in the World: Teaching and Learning the Language and Literatures; Papers of an International Conference Entitled 'Progress in English Studies' Held in London, 17 - 21 September 1984 to Celebrate the Fiftieth Anniversary of the British Council and Its Contribution to the Field of English Studies over Fifty Years*, edited by Randolph Quirk and Henry G. Widdowson, 212–23. Cambridge: Cambridge University Press.

———. 1990. *Genre Analysis: English in Academic and Research Settings.* Cambridge: Cambridge University Press.

Sweeting, Anthony. 2004. *Education in Hong Kong, 1941 to 2001: Visions and Revisions.* Hong Kong: Hong Kong University Press. http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10387592.

Sweeting, Anthony, and Edward Vickers. 2005. “On Colonizing 'Colonialism': The Discourses of the History of English in Hong Kong.” *World Englishes* 24 (2): 113–30. doi:10.1111/j.1467-971X.2005.00397.x.

Szczygłowska, Tatiana. 2013. “Various Shades of 'blue(s)': A Study of the Meaning Potential of the Colour Term 'blue' in the British Press.” *Linguistica Silesiana* 34: 297–318.

Szmrecsanyi, Benedikt, and Bernd Kortmann. 2011. “Typological Profiling: Learner Englishes versus Indigenized L2 Varieties of English.” In *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*, edited by Joybrato Mukherjee and Marianne Hundt, 167–87. Amsterdam: John Benjamins.

Taguchi, Naoko. 2002. “A Comparative Analysis of Discourse Markers in English Conversational Registers.” *Issues in Applied Linguistics* 13 (1): 41–69.

Tarde, Gabriel. (1898) 2010. “Opinion and Conversation.” In *Gabriel Tarde on Communication & Social Influence: Selected Papers*, edited by Terry Clark, 297–318. Chicago: University of Chicago Press.

Tay, Mary. 1991. “Southeast Asia and Hong Kong.” In *English around the World: Sociolinguistic Perspectives*, edited by Jenny Cheshire, 319–32. Cambridge: Cambridge University Press.

Taylor, Andrew. 1989. “Hong Kong’s English Newspapers: An Analysis of the Style and Content of the 'South China Morning Post' and the 'Hongkong Standard'.” *English Today* 5 (4): 18. doi:10.1017/S0266078400004302.

Taylor, Charlotte. 2010. “Science in the News: A Diachronic Perspective.” *Corpora* 5 (2): 221–50. doi:10.3366/cor.2010.0106.

Teich, Elke. 2003. *Cross-Linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts.* Berlin: De Gruyter Mouton.

Teich, Elke, and Mônica Holtz. 2009. “Scientific Registers in Contact: An Exploration of the Lexico-Grammatical Properties of Interdisciplinary Discourses.” *International Journal of Corpus Linguistics* 14 (4): 524–48. doi:10.1075/ijcl.14.4.04tei.

Teo, Peter. 2000. “Racism in the News: A Critical Discourse Analysis of News Reporting in Two Australian Newspapers.” *Discourse & Society* 11 (1): 7–49. doi:10.1177/0957926500011001002.

Tereszkiewicz, Anna. 2013. “Tweeting the News: A Contrastive Study of English and German Newspaper Tweets.” *Kwartalnik Neofilologiczny* 60 (3): 361–76.

The British National Corpus. 2007. 3rd ed. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Accessed February 28, 2016. http://www.natcorp.ox.ac.uk.

The Morning Star. 2015. “About Us.” Accessed November 18, 2015. http://www.morningstaronline.co.uk/about-us.

The Newspaper Works. 2015. “The Works Q3: Moving Target.” Accessed February 04, 2016. http://www.thenewspaperworks.com.au/the-works-q3-newspaper-audience-report/.

The Standard. 2015. “About Us.” Accessed January 26, 2016. http://www.thestandard.com.hk/about_us.php.

The State of the News Media. 2012. “Who Owns the Media: Top Newspaper Companies.” Accessed January 26, 2016. http://www.stateofthemedia.org/media-ownership/newspapers/.

The World Bank Group. 2012. “Data: Access to Electricity.” Accessed November 08, 2015. http://data.worldbank.org/indicator/EG.ELC.ACCS.ZS/countries/KE?display=graph.

Thiele, Wolfgang. 2000. “Textlinguistik im englischsprachigen Raum.” In *Text- Und Gesprächslinguistik/ Linguistics of Text and Conversation. Ein Internationales Handbuch Zeitgenössischer Forschung / an International Handbook of Contemporary Research*, edited by Klaus Brinker, Gerd Antos, Wolfgang Heinemann, and Sven Sager. 2 vols, 132–39 1. Berlin: De Gruyter Mouton.

Thomas, Richard. 1985. "Plain English and the Law." *Statute Law Review* 6 (1): 139–51. doi:10.1093/slr/6.1.139.

Thompson, Elizabeth A., Peter R. R. White, and Philip Kitley. 2008. "'Objectivity' and 'Hard News' Reporting across Cultures: Comparing the News Report in English, French, Japanese and Indonesian Journalism." *Journalism Studies* 9 (2): 212–28. doi:10.1080/14616700701848261.

Thompson, Geoff. 2004. *Introducing Functional Grammar.* London: Edward Arnold.

Thompson, Paul. 2005. "Points of Focus and Position: Intertextual Reference in PhD Theses." *Journal of English for Academic Purposes* 4 (4): 307–23. doi:10.1016/j.jeap.2005.07.006.

Thompson, Susan E. 2003. "Text-Structuring Metadiscourse, Intonation and the Signalling of Organisation in Academic Lectures." *Journal of English for Academic Purposes* 2 (1): 5–20. doi:10.1016/S1475-1585(02)00036-X.

Thurlow, Crispin. 2003. "Generation Txt? the Sociolinguistics of Young People's Text-Messaging." *Discourse Analysis Online* 1. http://extra.shu.ac.uk/daol/articles/v1/n1/a3/thurlow2002003.html. Accessed January 23, 2016.

Tsfati, Yariv. 2003. "Does Audience Skepticism of the Media Matter in Agenda Setting?" *Journal of Broadcasting & Electronic Media* 47 (2): 157–76. doi:10.1207/s15506878jobem4702_1.

Tucker, Jim. 2013. "Backrower Scott Higginbotham ready to answer Wallaby spring tour SOS." *The Courier Mail*, October 20. Accessed November 10, 2016. http://www.couriermail.com.au/sport/rugby/backrower-scott-higginbotham-ready-to-answer-wallaby-spring-tour-sos/story-fnii0mk9-1226743284355#.

Tucker, Paul. 2003. "Evaluation in the Art-Historical Research Article." *Journal of English for Academic Purposes* 2 (4): 291–312. doi:10.1016/S1475-1585(03)00047-X.

Turner, George W. 1972. "The Grammar of Newspaper Headlines Containing the Preposition on in the Sense 'about'." *Linguistics* 10 (87). doi:10.1515/ling.1972.10.87.71.

U.S. Census Bureau. 2015. "Population Clock." Accessed November 19, 2015. http://www.census.gov/popclock.

UK Audit Bureau of Circulations. 2015. "National Newspapers Report." Accessed November 14, 2015. http://abc.fileburst.com/publishedreports/National%20Newspapers%20Report%202015-09.pdf.

UK Office for National Statistics. 2013. "General Lifestyle Survey 2011." Accessed November 13, 2015. http://www.ons.gov.uk/ons/dcp171776_302199.pdf.

———. 2015a. "Internet Access - Households and Individuals 2015." Accessed November 13, 2015. http://www.ons.gov.uk/ons/dcp171778_412758.pdf.

———. 2015b. "Overview of the UK Population." Accessed November 07, 2015. http://www.ons.gov.uk/ons/dcp171776_422383.pdf.

UNESCO Institute for Statistics. 2015. "Education Literacy Rate." Accessed November 08, 2015. http://data.uis.unesco.org/index.aspx?queryname=166.

Ungerer, Friedrich. 1999. "Newspapers as Text Ensembles." In *English via Various Media*, edited by Hans-Jürgen Diller, Erwin Otto, and Gert Stratmann, 17–37. Anglistik und Englischunterricht 62. Heidelberg: Universitätsverlag C. Winter.

United Nations Statistics Division. 2016. "UNdata: Kenya." Accessed January 26, 2016. http://data.un.org/CountryProfile.aspx?crName=kenya.

Ure, Jean. 1971. "Lexical Density and Register Differentiation." In *Applications of Linguistics*, edited by George Perren and John Trim, 443–52. London: Cambridge University Press.

Vassileva, Irena. 2001. "Commitment and Detachment in English and Bulgarian Academic Writing." *English for Specific Purposes* 20 (1): 83–102. doi:10.1016/S0889-4906(99)00029-0.

Verschueren, Jef. 1989. "English as Object and Medium of (Mis)understanding." In *English across Cultures, Cultures across English*, edited by Ofelia García and Ricardo Otheguy, 31–53. Berlin: De Gruyter Mouton.

Vestergaard, Torben. 1999. "Free Adjuncts in Newspaper Discourse." In *English via Various Media*, edited by Hans-Jürgen Diller, Erwin Otto, and Gert Stratmann, 89–105. Anglistik und Englischunterricht 62. Heidelberg: Universitätsverlag C. Winter.

Wachira, Charles. 2014. "Kenyan Media Company Linked to Kenyatta Offers Free Daily Paper." *BloombergBusiness*, July 1. Accessed November 12, 2015. http://www.bloomberg.com/news/articles/2014-07-01/kenyan-media-company-linked-to-kenyatta-offers-free-daily-paper.

Wahl-Jorgensen, Karin. 2002. "The Construction of the Public in Letters to the Editor: Deliberative Democracy and the Idiom of Insanity." *Journalism* 3 (2): 183–204. doi:10.1177/146488490200300203.

———. 2010. "Letters to the Editor as a Forum for Public Deliberation: Modes of Publicity and Democratic Debate." *Critical Studies in Media Communication* 18 (3): 303–20. doi:10.1080/07393180128085.

Wang, Wei. 2008. "Newspaper Commentaries on Terrorism in China and Australia: A Contrastive Genre Study." In *Contrastive Rhetoric*, edited by Ulla Connor, Ed Nagelhout, and William Rozycki, 169–91. Pragmatics & Beyond New Series 169. Amsterdam: John Benjamins.

Warner, Anthony. 2005. "Why DO Dove: Evidence for Register Variation in Early Modern English." *Language Variation and Change* 17 (3): 257–80. doi:10.1017/S0954394505050106.

Weber, Max. 1911. "Geschäftsbericht der Deutschen Gesellschaft für Soziologie." In *Verhandlungen des Ersten Deutschen Soziologentages in Frankfurt a. M. 1910*, 39–62. Schriften der Deutschen Gesellschaft für Soziologie 1. Tübingen: Mohr.

———. (1921) 1976. *Soziologische Grundbegriffe.* 3rd ed. Tübingen: Mohr Siebeck.

Weeks, David. 2013. "What's got interim?" *The Sun*, March 1. Accessed November 10, 2016. http://www.thesun.co.uk/sol/homepage/features/4818700/Rafa-Benitez-Chelsea-rant-analysed-by-neuropsychologist.html.

Wegener, Rebekah. 2011. "Parameters of Context: From Theory to Model and Application." PhD-Thesis, Department of Linguistics, Macquarie University.

Welch, Michael, Melissa Fenwick, and Meredith Roberts. 1997. "Primary Definitions of Crime and Moral Panic: A Content Analysis of Experts' Quotes in Feature Newspaper Articles on Crime." *Journal of Research in Crime and Delinquency* 34 (4): 474–94. doi:10.1177/0022427897034004004.

———. 1998. "State Managers, Intellectuals, and the Media: A Content Analysis of Ideology in Experts' Quotes in Feature Newspaper Articles on Crime." *Justice Quarterly* 15 (2): 219–41. doi:10.1080/07418829800093721.

Wensing, Emma H., and Toni Bruce. 2003. "Bending the Rules: Media Representations of Gender During an International Sporting Event." *International Review for the Sociology of Sport* 38 (4): 387–96. doi:10.1177/1012690203384001.

Wenzel, Peter. 2015. "Endings in Literature: A Survey." In *Last Things: Essays on Ends and Endings*, edited by Gavin Hopps, Stella Neumann, Sven Strasen, and Peter Wenzel, 19–34. Aachen British and American Studies 19. Frankfurt am Main: Peter Lang.

Werlich, Egon. 1975. *Typologie der Texte.* Heidelberg: UTB.

———. 1976. *A Text Grammar of English.* Heidelberg: UTB.

Westin, Ingrid. 2002. *Language Change in English Newspaper Editorials.* New York: Rodopi.

White, Peter R. R. 1997. "Death, Disruption and the Moral Order: The Narrative Impulse in Mass Media 'Hard News' Reporting." In *Genre and Institutions. Social Processes in the Workplace and School*, edited by James R. Martin and Frances Christie, 101–33. London: Cassell.

———. 2000. "Media Objectivity and the Rhetoric of News Story Structure." In *Discourse and Community: Doing Functional Linguistics*, edited by Eija Ventola, 379–97. Language Performance 21. Tübingen: Narr.

———. 2004. "Subjectivity, Evaluation and Point of View in Media Discourse." In *Applying English Grammar: Functional and Corpus Approaches*, edited by Caroline Coffin and Kieran O'Halloran, 229–46. London, New York: Edward Arnold.

———. 2006. "Evaluative Semantics and Ideological Positioning in Journalistic Discourse: a New Framework for Analysis." In *Mediating Ideology in Text and Image: Ten Critical Studies*, edited by Inger Lassen, Jeanne Strunck, and Torben Vestergaard, 37–67. Amsterdam: John Benjamins.

Widdowson, Henry G. 1994. "The Ownership of English." *TESOL Quarterly* 28 (2): 377–89. doi:10.2307/3587438.

———. 1997. "EIL, ESL, EFL: Global Issues and Local Interests." *World Englishes* 16 (1): 135–46. doi:10.1111/1467-971X.00054.

———. 2004. *Text, Context, Pretext. Critical Issues in Discourse Analysis.* Language in Society 35. Malden: Blackwell.

Willnat, Lars, and Jian-Hua Zhu. 1996. "Newspaper Coverage and Public Opinion in Hong Kong: A Time-Series Analysis of Media Priming." *Political Communication* 13 (2): 231–46. doi:10.1080/10584609.1996.9963109.

Wimmel, Andreas. 2009. "Beyond the Bosphorus? Comparing Public Discourses on Turkey's EU Application in the German, French and British Quality Press." *Journal of Language and Politics* 8 (2): 223–43. doi:10.1075/jlp.8.2.03wim.

Wodak, Ruth, and Michael Meyer. 2009. "Critical Discourse Analysis: History, Agenda, Theory and Methodology." In *Methods of Critical Discourse Analysis*, edited by Ruth Wodak and Michael Meyer. 2nd ed., 1–33. London: Sage Publications.

Wong, May L.-Y. 2010. "Expressions of Gratitude by Hong Kong Speakers of English: Research from the International Corpus of English in Hong Kong (ICE-HK)." *Journal of Pragmatics* 42 (5): 1243–57. doi:10.1016/j.pragma.2009.09.022.

———. 2014. "Verb-Preposition Constructions in Hong Kong English: A Cognitive Semantic Account." *Linguistics* 52 (3): 603–35. doi:10.1515/ling-2014-0001.

World Health Organisation. 2000. "Preventing Suicide: A Resource for Media Professionals." Accessed December 11, 2015. http://www.who.int/mental_health/media/en/426.pdf.

Xiao, Richard. 2009. "Multidimensional Analysis and the Study of World Englishes." *World Englishes* 28 (4): 421–50. doi:10.1111/j.1467-971X.2009.01606.x.

Xiao, Zhonghua, and Tony McEnery. 2005. "Two Approaches to Genre Analysis: Three Genres in Modern American English." *Journal of English Linguistics* 33 (1): 62–82. doi:10.1177/0075424204273957.

Yang, Jin. 2003. "Framing the NATO Air Strikes on Kosovo across Countries: Comparison of Chinese and US Newspaper Coverage." *International Communication Gazette* 65 (3): 231–49. doi:10.1177/0016549203065003002.

Yao, Xinhue, and Peter Collins. 2012. "The Present Perfect in World Englishes." *World Englishes* 31 (3): 386–403. doi:10.1111/j.1467-971X.2012.01756.x.

Yates, Simeon J. 1996. “Oral and Written Linguistic Aspects of Computer Conferencing.” In *Computer-Mediated Communication*, edited by Susan C. Herring, 29–46. Pragmatics & Beyond New Series 39. Amsterdam: John Benjamins.

Yip, Paul S. F., King-Wa Fu, Kris C. T. Yang, Brian Y. T. Ip, Cecilia L. W. Chan, Eric Y. H. Chen, Dominic T. S. Lee, Frances Y. W. Law, and Keith Hawton. 2006. “The Effects of a Celebrity Suicide on Suicide Rates in Hong Kong.” *Journal of Affective Disorders* 93 (1-3): 245–52. doi:10.1016/j.jad.2006.03.015.

Yoshikawa, Hiroshi. 2005. “Recognition of World Englishes: Changes in Chukyo University Students' Attitudes.” *World Englishes* 24 (3): 351–60. doi:10.1111/j.0083-2919.2005.00416.x.

Yung, Vicki K.-y. 1996. “The Presentation of Voice in Chinese and English Newspapers in Hong Kong.” *Perspectives, Working Papers of the Department of English* 8 (1): 64–96.

Zappavigna, Michele. 2011. “Ambient Affiliation: A Linguistic Perspective on Twitter.” *New Media & Society* 13 (5): 788–806. doi:10.1177/1461444810385097.

Zhou, Xiang. 2008. “Cultural Dimensions and Framing the Internet in China: A Cross-Cultural Study of Newspapers' Coverage in Hong Kong, Singapore, the US and the UK.” *International Communication Gazette* 70 (2): 117–36. doi:10.1177/1748048507086908.

Zhu, Yunxia, and Bernard McKenna. 2012. “Legitimating a Chinese Takeover of an Australian Iconic Firm: Revisiting Models of Media Discourse of Legitimacy.” *Discourse & Society* 23 (5): 525–52. doi:10.1177/0957926512452971.

Zillmann, Dolf. 1979. *Hostility and Aggression.* Hillsdale: Lawrence Erlbaum.

Zuengler, Jane E. 1983. “Kenyan English.” In *The Other Tongue: English across Cultures*, edited by Braj B. Kachru. 1st ed., 112–24. Oxford: Pergamon Press.

Appendix 1: Part of Speech-Tags

Part of speech-tag	**Corresponding parts of speech**
CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential *there*
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or mass
NNS	Noun, plural
NP	Proper noun, singular
NPS	Proper noun, plural
PDT	Predeterminer
POS	Possessive ending
PP	Personal pronoun
PP$	Possessive pronoun
RB	Adverb
RBF	Adverb (non-lexical)
RBR	Adverb, comparative
RBFR	Adverb, comparative (non-lexical)
RBS	Adverb, superlative
RBFS	Adverb, superlative (non-lexical)
RP	Particle
SYM	Symbol
TO	*to*
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund or present participle
VBN	Verb, past participle
VBP	Verb, non-3rd person singular present
VBZ	Verb, 3rd person singular present
WDT	Wh-determiner
WP	Wh-pronoun
WP$	Possessive wh-pronoun
WRB	Wh-adverb
SENT	. ? !
,	,
:	- : ; ...
``	Opening quotation marks
''	Closing quotation marks
(	Opening bracket
)	Closing bracket

(cf. Santorini 1991)

Appendix 2: Most Frequent Lexical Items

Economy														
Australia			Hong Kong			Kenya			UK			USA		
Lemma	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**
say	773	1.09	say	986	1.20	say	667	1.02	say	557	0.66	say	825	0.97
cent	582	0.82	China	609	0.74	Kenya	418	0.64	year	407	0.48	year	408	0.48
year	382	0.54	year	557	0.68	cent	281	0.43	bank	357	0.42	company	347	0.41
market	294	0.41	market	493	0.60	year	260	0.40	cent	275	0.33	stock	297	0.35
company	256	0.36	company	342	0.42	government	247	0.38	market	270	0.32	market	286	0.34
business	255	0.36	billion	284	0.35	billion	235	0.36	share	228	0.27	bank	277	0.32
Mr	243	0.34	new	243	0.30	market	214	0.33	do	227	0.27	million	260	0.31
million	233	0.33	fund	237	0.29	bank	187	0.28	company	218	0.26	billion	248	0.29
price	206	0.29	percent	230	0.28	company	169	0.26	make	218	0.26	new	243	0.29
rate	206	0.29	bank	223	0.27	country	168	0.26	last	208	0.25	do	229	0.27

Hard News														
Australia			Hong Kong			Kenya			UK			USA		
Lemma	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**
say	719	1.31	say	908	1.40	say	778	1.33	say	741	0.90	say	1,446	1.70
police	344	0.63	police	409	0.63	police	479	0.82	police	441	0.53	police	463	0.54
man	267	0.49	Mr	292	0.45	officer	194	0.33	do	284	0.34	Mr	322	0.38
Mr	213	0.39	year	238	0.37	court	187	0.32	year	281	0.34	do	248	0.29
court	165	0.30	people	191	0.29	county	176	0.30	Mr	231	0.28	man	214	0.25
do	146	0.27	report	179	0.28	suspect	167	0.29	crime	201	0.24	year	201	0.24
charge	121	0.22	do	161	0.25	take	155	0.27	tell	191	0.23	charge	186	0.22
year	116	0.21	court	148	0.23	kill	146	0.25	man	181	0.22	people	184	0.22
people	115	0.21	case	139	0.21	people	144	0.25	take	180	0.22	find	182	0.21
crime	113	0.21	man	132	0.20	body	114	0.20	find	178	0.22	case	174	0.20

Lifestyle														
Australia			Hong Kong			Kenya			UK			USA		
Lemma	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**
say	895	0.88	say	1,015	0.86	say	862	0.68	do	650	0.57	say	1,132	0.96
do	587	0.58	do	422	0.36	do	561	0.44	say	589	0.52	do	546	0.46

people	290	0.29	year	366	0.31	year	335	0.26	make	317	0.28	new	360	0.31
time	253	0.25	make	279	0.24	make	330	0.26	people	317	0.28	get	328	0.28
make	249	0.25	people	271	0.23	woman	313	0.25	get	298	0.26	make	320	0.27
year	247	0.24	also	270	0.23	child	299	0.24	year	264	0.23	year	292	0.25
go	237	0.23	Hong	261	0.22	use	292	0.23	go	258	0.23	people	248	0.21
also	217	0.21	new	258	0.22	get	286	0.23	time	251	0.22	go	235	0.20
take	213	0.21	Kong	254	0.22	time	282	0.22	woman	245	0.21	time	232	0.20
use	195	0.19	use	227	0.19	take	271	0.21	child	232	0.20	woman	220	0.19
Politics														
Australia			Hong Kong			Kenya			UK			USA		
Lemma	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**
say	1,081	1.41	say	1,101	1.34	say	982	1.31	say	711	0.80	say	1,053	1.18
Mr	733	0.96	China	449	0.55	president	316	0.42	Mr	436	0.49	Mr	468	0.53
government	462	0.60	Mr	395	0.48	county	313	0.42	do	327	0.37	do	345	0.39
minister	351	0.46	government	393	0.48	Mr	305	0.41	party	245	0.28	president	299	0.34
labor	337	0.44	party	350	0.43	government	299	0.40	year	231	0.26	new	246	0.28
election	286	0.37	minister	332	0.40	MP	290	0.39	minister	224	0.25	year	246	0.28
do	274	0.36	year	251	0.30	party	275	0.37	make	221	0.25	Obama	240	0.27
Abbott	233	0.30	also	213	0.26	court	257	0.34	labour	207	0.23	government	239	0.27
year	227	0.30	election	213	0.26	Kenya	256	0.34	last	193	0.22	party	229	0.26
Ms	216	0.28	country	192	0.23	election	233	0.31	country	191	0.22	vote	208	0.23
Sports														
Australia			Hong Kong			Kenya			UK			USA		
Lemma	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**	**Lemma**	**Freq.**	**%**
say	679	0.85	say	822	0.95	say	429	0.66	say	527	0.53	say	806	0.82
do	338	0.42	team	377	0.44	Kenya	416	0.64	do	526	0.53	do	525	0.54
play	311	0.39	sport	365	0.42	team	370	0.57	go	357	0.36	game	459	0.47
game	293	0.37	year	353	0.41	win	290	0.45	get	326	0.33	play	384	0.39
get	268	0.33	Mr	303	0.35	match	231	0.36	win	309	0.31	team	375	0.38
year	258	0.32	do	297	0.34	club	214	0.33	team	305	0.31	win	363	0.37
player	254	0.32	win	290	0.34	year	193	0.30	play	295	0.30	get	353	0.36
win	252	0.31	game	276	0.32	player	191	0.30	England	289	0.29	go	324	0.33
go	251	0.31	player	274	0.32	league	180	0.28	year	271	0.27	year	296	0.30
good	234	0.29	world	248	0.29	world	173	0.27	time	267	0.27	first	285	0.29

Appendix 3: Top Ten Keywords

Domain-internal

Economy

Australia			Hong Kong			Kenya			UK			USA		
Word	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**
Australian	168	459	China	625	1,173	Sh	446	1,523	bn	160	431	U	229	317
Australia	176	435	Yuan	226	655	Kenya	420	1,492	UK	204	422	S	1,344	184
cent	552	341	HK	200	568	Africa	141	395	Britain	124	274	Gazprom	46	124
per	556	307	percent	235	380	Nairobi	86	305	pc	100	194	Dow	74	119
US	292	218	mainland	129	348	African	94	290	p	169	179	CEO	81	113
WA	51	157	Kong	181	338	county	82	231	London	82	125	Apple	87	106
iron	64	152	Hong	176	325	Kenyan	60	213	chancellor	49	114	stocks	116	103
Queensland	44	149	Li	88	254	farmers	71	158	Lloyds	40	112	wall	80	97
mining	65	133	Chinese	170	233	sugar	54	155	British	58	99	street	95	91
Mr	243	125	Beijing	93	219	Kenyans	40	142	FTSE	52	96	Netflix	29	88

Hard News

Australia			Hong Kong			Kenya			UK			USA		
Word	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**
Brisbane	48	176	Kong	114	362	Kenya	100	344	her	559	180	Boston	97	198
Australian	45	139	China	119	352	Nairobi	96	330	miss	91	161	Dutschke	68	190
man	208	128	Hong	111	352	Sh	84	298	British	92	151	Curtis	64	179
Melbourne	44	123	Chinese	89	255	county	169	281	she	474	150	attorney	88	151
Ms	109	118	Beijing	57	190	Mombasa	77	273	Philpott	50	143	Texas	54	141
Perth	30	110	HK	57	190	the	4,415	237	London	76	138	Dorner	48	134
NSW	29	106	Yuan	45	150	Boda	42	149	cannabis	50	134	FBI	58	108
Queensland	29	106	province	55	139	hawkers	41	145	Berezovsky	46	131	York	65	105
robbery	94	102	percent	43	134	investigations	68	145	I	434	121	snow	41	99
armed	77	95	Chen	39	130	land	60	119	Manchester	38	108	federal	74	99

Lifestyle

Australia			Hong Kong			Kenya			UK			USA		
Word	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**
Australia	158	417	Kong	262	756	Kenya	243	716	Marina	42	136	s	1,623	203
Australian	120	325	Hong	261	753	Sh	213	635	I	1,158	123	York	127	193

Sydney	51	150	China	186	403	Nairobi	116	307	monkeys	33	106	Manhattan	39	99
you	824	150	Chinese	171	361	Moi	74	225	Britain	54	97	brain	106	92
your	416	131	HK	93	295	he	986	195	Conny	25	81	acne	41	90
WA	58	128	Beijing	85	244	him	249	173	garden	68	80	new	358	84
skin	124	108	Romney	61	193	African	90	163	UK	60	80	breast	59	78
au	40	105	Yuan	53	168	Kenyans	48	145	sugar	67	75	game	62	74
Perth	28	97	mainland	54	135	was	1,023	145	Ulrike	23	74	a	3,469	71
Melbourne	32	92	Shanghai	47	133	county	68	141	we	620	71	apnea	22	70
Politics														
Australia			Hong Kong			Kenya			UK			USA		
Word	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**
labor	340	800	China	482	1,002	Kenya	262	799	labour	211	504	Berlusconi	147	282
Abbott	233	645	Hong	156	480	county	234	652	Blair	141	385	Obama	245	279
Gillard	214	533	Chinese	202	452	Ruto	187	608	Britain	160	300	gun	143	275
Rudd	211	483	Kong	151	419	ICC	181	580	Cameron	174	275	sen	78	220
Australia	201	426	Leung	93	298	Uhuru	159	521	Brown	88	151	Weiner	69	210
Queensland	117	394	Beijing	116	274	cord	148	504	EU	171	148	Quinn	59	180
Australian	139	326	Japan	139	239	jubilee	130	443	UK	80	140	s	1,441	159
WA	75	222	Li	100	231	ODM	124	422	Lib	54	134	Italy	103	158
Mr	733	220	HK	68	218	Sh	123	419	Tory	48	126	mediaset	54	155
federal	128	208	Chan	48	154	Kenyatta	146	372	Russia	76	125	senate	147	135
Sports (see also Fest, forthcoming)														
Australia			Hong Kong			Kenya			UK			USA		
Word	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**
Sydney	99	305	Mr	303	697	Kenya	418	1,424	England	281	379	Knicks	95	263
Australian	161	263	China	248	670	Gor	131	497	Hamilton	81	200	Giants	108	240
Australia	170	257	Chinese	159	448	Nairobi	125	462	I	1,166	167	Yankees	101	214
Melbourne	123	246	Hong	137	404	Sh	114	432	Froome	49	142	Boston	106	199
wallabies	66	197	Kong	132	395	Leopards	81	307	Capello	50	119	Mets	65	192
Perth	61	194	sport	245	282	Mahia	81	307	Britain	59	110	playoff	95	191
Brisbane	64	190	Li	71	197	sevens	77	261	Mclaren	52	108	Jeter	67	188
AF	70	161	sumo	60	192	Thika	56	212	Wiggins	37	107	playoffs	67	164
blacks	43	134	soccer	107	190	marathon	107	207	his	963	102	Harvey	71	155
WA	42	131	Asian	61	149	Sofapaka	54	204	Mercedes	42	97	Cruz	56	149

Variety-internal

Australia

Economy			Hard News			Lifestyle			Politics			Sports[29]		
Word	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**
per	556	843	police	341	1,071	you	824	724	labor	340	956	game	246	655
cent	522	827	man	208	481	your	416	658	Mr	733	900	players	185	515
market	247	656	robbery	94	366	can	459	275	minister	328	843	coach	155	472
its	357	383	court	166	360	skin	124	264	election	286	769	season	197	459
company	196	367	was	829	356	says	350	263	government	445	733	league	122	330
business	199	312	armed	77	299	or	462	194	Rudd	211	658	club	147	297
bank	124	275	crime	94	299	are	738	188	Gillard	214	637	his	598	288
prices	103	273	charged	88	277	women	137	187	Abbott	233	611	win	149	276
us	292	264	murder	72	250	children	150	187	prime	185	511	rugby	84	262
dollar	101	259	crash	69	219	exercise	64	141	coalition	139	348	team	161	254

Hong Kong

Economy			Hard News			Lifestyle			Politics			Sports		
Word	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**
market	462	845	police	404	1,118	says	647	779	minister	307	822	sport	245	677
billion	284	505	fire	112	310	you	441	411	party	319	684	team	303	666
cent	211	399	crime	91	288	fashion	113	261	election	213	600	game	199	572
per	247	369	were	346	243	your	154	204	government	395	445	players	198	493
trading	139	369	killed	80	222	I	611	194	political	181	398	games	179	493
Yuan	226	345	court	132	197	style	116	188	prime	136	309	sports	171	382
funds	135	337	arrested	73	190	can	391	162	vote	76	211	league	161	351
China	625	323	was	594	183	music	67	157	democratic	74	184	golf	119	338
investors	150	293	accident	58	182	like	241	155	votes	50	165	tournament	91	293
percent	235	279	trial	62	158	she	373	152	Rudd	47	155	cup	105	280

Kenya

Economy			Hard News			Lifestyle			Politics			Sports		
Word	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**
billion	233	522	police	479	1,256	I	1,042	655	election	232	615	team	275	659
per	328	502	were	433	331	you	563	603	party	229	573	league	183	614

[29] For keywords in sports news, see also Fest (forthcoming).

cent	281	469	officers	123	310	your	289	485	president	312	514	club	175	504
Sh	446	442	killed	99	259	her	486	333	ICC	181	511	match	155	476
tax	125	353	arrested	78	209	says	452	329	cord	148	467	coach	135	466
market	154	307	suspects	75	202	my	391	328	Ruto	187	430	Gor	131	460
financial	128	299	accident	64	191	she	519	316	ODM	124	381	cup	123	402
bank	144	292	said	680	191	hair	111	230	Uhuru	159	358	players	145	379
shares	85	244	investigations	68	181	child	132	196	MP	167	342	marathon	107	374
sector	115	242	fire	82	179	it	1,064	193	jubilee	130	330	win	143	351
UK														
Economy			Hard News			Lifestyle			Politics			Sports		
Word	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**
bank	250	664	police	430	1,038	you	588	312	Cameron	174	481	team	310	630
its	449	528	was	1,299	447	my	399	280	labour	211	469	England	281	547
shares	157	484	her	559	420	your	220	231	Mr	436	449	players	165	374
market	210	484	crime	155	378	I	1,158	197	election	172	445	cup	150	366
bn	160	467	murder	111	304	women	192	182	Obama	137	428	game	186	366
p	167	439	officers	97	271	she	568	176	Blair	141	416	I	1,166	354
stock	137	383	death	153	260	sugar	67	172	minister	193	411	season	154	340
per	305	370	accident	92	250	food	124	169	party	211	396	match	117	279
cent	275	342	court	152	235	hair	76	147	EU	171	370	his	963	278
investors	104	324	robbery	63	219	garden	68	141	prime	141	311	he	1,249	264
USA														
Economy			Hard News			Lifestyle			Politics			Sports		
Word	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**	**Word**	**Freq.**	**Key.**
stock	197	602	police	464	1,250	says	622	633	Obama	245	530	season	300	695
billion	245	589	said	1,268	523	your	297	407	President	293	493	game	333	681
investors	198	553	murder	130	377	you	710	403	Berlusconi	147	477	team	290	581
market	221	441	was	1,009	370	brain	106	227	party	199	457	play	192	389
bank	182	404	were	433	272	women	169	204	election	158	434	I	831	330
its	444	385	authorities	104	272	researchers	70	194	senate	147	424	win	185	320
stocks	116	383	death	115	255	sleep	84	187	Mr	469	402	Yankees	101	308
shares	129	380	arrested	97	254	or	490	184	political	174	375	Knicks	95	300
index	115	380	officers	80	243	art	79	184	Merkel	105	311	Giants	108	292
company	222	351	Dutschke	68	233	cancer	77	169	vote	125	308	playoff	95	282

Appendix 4: Ranges and Deviations for All Features

Australia													
Linguistic Feature	**Values**							**Deviations from average**					
	ECO	HN	LIFE	POL	SPO	Ø	Range	ECO	HN	LIFE	POL	SPO	|∑|
Lexical Density	60.10	60.35	58.98	61.03	58.00	59.69	3.03	0.32	0.07	1.44	-0.61	2.42	4.86
Nouns	24.15	22.39	22.25	21.71	18.03	21.71	6.12	-2.44	-0.68	-0.54	0.00	3.68	7.35
Past Tense	50.00	72.09	27.33	48.07	47.38	48.97	44.76	-1.03	-23.12	21.64	0.91	1.60	48.29
Declarative Mood	98.84	98.92	93.50	98.71	98.68	97.73	5.42	-1.11	-1.19	4.23	-0.98	-0.95	8.46
Personal Pronouns	3.43	5.17	6.75	4.41	6.62	5.28	3.31	1.84	0.11	-1.47	0.87	-1.34	5.63
Type-token-ratio	44.40	46.49	40.98	42.88	42.61	43.47	5.51	-0.93	-3.02	2.49	0.59	0.86	7.9
Sentence-initial coor-dinators	0.17	0.07	0.20	0.18	0.17	0.16	0.13	-0.01	0.09	-0.04	-0.02	-0.02	0.18
Contractions	0.51	0.48	1.45	0.52	1.67	0.93	1.19	0.42	0.45	-0.52	0.41	-0.74	2.54
Mean word length	4.85	4.71	4.64	4.85	4.51	4.71	0.34	-0.14	0.00	0.07	-0.14	0.20	0.55
Modality	1.10	0.88	1.35	1.67	1.41	1.28	0.79	0.18	0.40	-0.07	-0.39	-0.13	1.17
Imperative Mood	0.56	0.81	3.67	0.59	0.43	1.21	3.24	0.65	0.40	-2.46	0.63	0.78	4.92
Interrogative Mood	0.60	0.27	2.83	0.71	0.89	1.06	2.56	0.46	0.79	-1.77	0.35	0.17	3.54
Pers. Pronouns (1st & 2nd person)	0.82	0.98	3.11	1.57	2.42	1.78	2.29	0.96	0.80	-1.33	0.21	-0.64	3.94
Nominalisations	2.80	1.56	1.76	3.41	1.19	2.15	2.22	-0.66	0.58	0.38	-1.27	0.96	3.85
Passive Voice	10.07	18.39	8.20	12.37	8.72	11.55	10.19	1.48	-6.84	3.35	-0.82	2.83	15.31
Boosters / Minimisers	0.84	0.49	1.06	0.64	1.02	0.81	0.57	-0.03	0.32	-0.25	0.17	-0.21	0.98
Titles	0.45	0.72	0.23	1.78	0.03	0.64	1.75	0.19	-0.08	0.41	-1.13	0.62	2.43
Eval. Adverbs / Adjec-tives	0.78	0.29	0.83	0.48	0.98	0.67	0.69	-0.11	0.38	-0.16	0.19	-0.31	1.15
Article Length	444.16	342.39	633.48	478.83	501.06	479.98	291.09	35.82	137.60	-153.49	1.16	-21.08	349.15

Hong Kong													
Linguistic Feature	**Values**							**Deviations from average**					
	ECO	HN	LIFE	POL	SPO	Ø	Range	ECO	HN	LIFE	POL	SPO	|∑|
Lexical Density	62.64	61.22	60.57	61.71	59.14	61.06	3.50	-1.58	-0.16	0.49	-0.65	1.92	4.8
Nouns	25.33	23.93	23.88	22.63	20.27	23.21	5.06	-2.12	-0.72	-0.67	0.58	2.94	7.03
Past Tense	41.67	72.08	26.71	51.51	40.00	46.39	45.36	4.73	-25.68	19.68	-5.11	6.39	61.6
Declarative Mood	99.47	99.46	97.02	99.22	98.84	98.80	2.46	-0.67	-0.66	1.79	-0.42	-0.04	3.57
Personal Pronouns	2.37	3.17	4.92	3.31	4.94	3.74	2.57	1.37	0.57	-1.18	0.43	-1.20	4.75
Type-token-ratio	42.30	45.77	40.26	43.19	42.79	42.86	5.52	0.56	-2.91	2.60	-0.33	0.07	6.48
Sentence-initial coordinators	0.13	0.10	0.20	0.12	0.16	0.14	0.11	0.01	0.05	-0.06	0.02	-0.02	0.16
Contractions	0.25	0.28	0.73	0.22	0.85	0.47	0.63	0.22	0.19	-0.26	0.25	-0.38	1.3
Mean word length	5.02	4.90	4.80	5.07	4.66	4.89	0.41	-0.13	-0.01	0.09	-0.18	0.23	0.64
Modality	1.11	0.83	1.04	1.31	1.15	1.09	0.47	-0.02	0.25	0.04	-0.22	-0.06	0.6
Imperative Mood	0.29	0.10	1.37	0.09	0.40	0.45	1.28	0.16	0.35	-0.92	0.36	0.05	1.83
Interrogative Mood	0.23	0.44	1.62	0.69	0.77	0.75	1.38	0.51	0.31	-0.87	0.06	-0.02	1.77
Pers. Pronouns (1st & 2nd person)	0.34	0.38	1.63	0.49	1.44	0.86	1.29	0.52	0.48	-0.77	0.37	-0.58	2.72
Nominalisations	3.09	2.45	2.01	3.99	1.81	2.67	2.18	-0.43	0.22	0.66	-1.32	0.86	3.49
Passive Voice	8.75	16.92	8.59	12.02	8.61	10.98	8.33	2.23	-5.94	2.39	-1.04	2.37	13.97
Boosters / Minimisers	0.67	0.59	0.92	0.61	0.88	0.73	0.33	0.06	0.15	-0.18	0.12	-0.15	0.66
Titles	0.19	0.60	0.30	1.06	0.45	0.52	0.88	0.33	-0.08	0.22	-0.54	0.07	1.24
Eval. Adverbs / Adjectives	0.57	0.39	0.68	0.52	0.78	0.59	0.39	0.02	0.20	-0.09	0.07	-0.19	0.57
Article Length	512.28	405.95	735.01	514.38	539.78	541.48	329.06	29.20	135.53	-193.54	27.10	1.70	387.07

Kenya													
Linguistic Feature	**Values**							**Deviations from average**					
	ECO	HN	LIFE	POL	SPO	Ø	Range	ECO	HN	LIFE	POL	SPO	|∑|
Lexical Density	61.06	60.38	57.55	60.34	60.90	60.05	3.51	-1.014	-0.334	2.496	-0.294	-0.854	4.99
Nouns	24.66	23.11	22.05	20.21	19.37	21.88	5.29	-2.78	-1.23	-0.17	1.67	2.51	8.36
Past Tense	40.00	75.00	32.22	53.95	50.26	50.29	42.78	10.29	-24.71	18.06	-3.66	0.02	56.75
Declarative Mood	99.28	98.97	95.51	98.44	98.35	98.11	3.77	-1.17	-0.86	2.60	-0.33	-0.24	5.19
Personal Pronouns	2.76	4.72	7.07	4.17	4.77	4.70	4.30	1.94	-0.02	-2.37	0.53	-0.07	4.93
Type-token-ratio	43.13	45.39	37.69	41.87	43.39	42.29	7.69	-0.84	-3.09	4.60	0.42	-1.10	10.05
Sentence-initial coordinators	0.05	0.06	0.19	0.10	0.12	0.10	0.14	0.05	0.04	-0.08	0.00	-0.02	0.2
Contractions	0.08	0.06	0.29	0.07	0.22	0.14	0.23	0.06	0.08	-0.15	0.07	-0.08	0.44
Mean word length	5.05	4.87	4.67	4.95	4.74	4.86	0.38	-0.19	-0.01	0.19	-0.09	0.12	0.6
Modality	1.12	0.81	1.20	1.31	1.20	1.13	0.50	0.01	0.32	-0.07	-0.18	-0.07	0.66
Imperative Mood	0.25	0.74	1.95	0.45	0.78	0.83	1.70	0.58	0.10	-1.12	0.39	0.06	2.23
Interrogative Mood	0.47	0.29	2.54	1.12	0.88	1.06	2.25	0.59	0.76	-1.48	-0.06	0.18	3.07
Pers. Pronouns (1st & 2nd person)	0.71	0.78	2.62	0.87	1.59	1.31	1.91	0.60	0.53	-1.31	0.44	-0.28	3.16
Nominalisations	3.89	2.57	2.29	3.53	1.41	2.74	2.48	-1.16	0.17	0.45	-0.79	1.33	3.9
Passive Voice	13.96	20.19	11.42	14.83	9.91	14.06	10.28	0.11	-6.13	2.65	-0.77	4.15	13.8
Boosters / Minimisers	0.54	0.36	0.78	0.35	0.49	0.51	0.43	-0.04	0.14	-0.27	0.15	0.02	0.62
Titles	0.28	0.35	0.24	1.08	0.06	0.40	1.02	0.12	0.05	0.16	-0.68	0.34	1.36
Eval. Adverbs / Adjectives	0.51	0.29	0.61	0.36	0.73	0.50	0.44	-0.01	0.21	-0.11	0.14	-0.23	0.7
Article Length	410.61	364.71	790.1	468.64	404.6	487.73	425.39	77.12	123.02	-302.37	19.09	83.13	604.73

UK													
Linguistic Feature	**Values**							**Deviations from average**					
	ECO	HN	LIFE	POL	SPO	Ø	Range	ECO	HN	LIFE	POL	SPO	|∑|
Lexical Density	60.16	58.75	57.58	59.34	56.62	58.49	3.54	-1.67	-0.26	0.91	-0.85	1.87	5.56
Nouns	23.83	21.38	21.38	20.94	17.02	20.91	6.81	-2.92	-0.47	-0.47	-0.03	3.89	7.78
Past Tense	42.21	68.93	35.24	44.91	42.86	46.83	33.69	4.62	-22.10	11.59	1.92	3.97	44.21
Declarative Mood	97.42	98.72	93.82	97.73	97.67	97.07	4.90	-0.35	-1.65	3.25	-0.66	-0.60	6.50
Personal Pronouns	3.61	6.54	7.42	5.15	7.25	5.99	3.81	2.38	-0.54	-1.43	0.84	-1.25	6.44
Type-token-ratio	43.92	42.47	40.19	43.98	41.65	42.44	3.79	-1.48	-0.03	2.25	-1.54	0.79	6.09
Sentence-initial coordinators	0.27	0.17	0.33	0.21	0.31	0.26	0.16	-0.01	0.09	-0.08	0.05	-0.05	0.28
Contractions	0.33	0.46	1.20	0.50	1.16	0.73	0.87	0.40	0.27	-0.47	0.23	-0.43	1.8
Mean word length	4.77	4.64	4.55	4.77	4.44	4.63	0.33	-0.14	-0.01	0.08	-0.14	0.19	0.56
Modality	1.37	0.92	1.34	1.38	1.43	1.29	0.51	-0.08	0.37	-0.05	-0.09	-0.14	0.73
Imperative Mood	0.68	0.28	1.58	0.41	0.45	0.68	1.30	0.00	0.40	-0.90	0.27	0.23	1.81
Interrogative Mood	1.90	1.05	4.59	1.85	1.88	2.25	3.55	0.36	1.21	-2.34	0.40	0.37	4.68
Pers. Pronouns (1st & 2nd person)	0.80	1.32	3.18	1.36	2.57	1.85	2.38	1.05	0.53	-1.33	0.49	-0.72	4.12
Nominalisations	2.25	1.53	1.48	2.56	1.19	1.80	1.37	-0.45	0.27	0.32	-0.76	0.62	2.41
Passive Voice	10.50	19.01	9.58	11.77	8.12	11.80	10.89	1.29	-7.21	2.22	0.03	3.68	14.43
Boosters / Minimisers	0.81	0.63	1.08	0.68	0.93	0.83	0.46	0.01	0.20	-0.26	0.15	-0.11	0.72
Titles	0.36	0.57	0.13	1.19	0.06	0.46	1.13	0.10	-0.11	0.34	-0.73	0.40	1.68
Eval. Adverbs / Adjectives	0.73	0.48	0.76	0.52	0.92	0.68	0.44	-0.05	0.20	-0.08	0.16	-0.24	0.73
Article Length	528.41	516.15	713.33	553.41	624.34	587.13	197.18	58.72	70.98	-126.20	33.72	-37.21	326.83

USA													
Linguistic Feature	**Values**							**Deviations from average**					
	ECO	HN	LIFE	POL	SPO	Ø	Range	ECO	HN	LIFE	POL	SPO	|∑|
Lexical Density	61.37	60.90	60.09	61.22	57.26	60.17	4.11	-1.20	-0.73	0.08	-1.05	2.91	5.97
Nouns	24.36	22.54	22.69	21.49	17.78	21.77	6.58	-2.59	-0.77	-0.92	0.28	3.99	8.55
Past Tense	50.00	73.61	27.53	49.64	43.67	48.89	46.08	-1.11	-24.72	21.36	-0.75	5.22	53.16
Declarative Mood	98.57	98.86	94.49	98.42	97.69	97.61	4.37	-0.97	-1.25	3.12	-0.81	-0.09	6.24
Personal Pronouns	3.13	4.95	5.72	4.36	6.49	4.93	3.36	1.80	-0.02	-0.79	0.57	-1.56	4.75
Type-token-ratio	42.55	43.06	39.46	45.16	39.08	41.86	6.08	-0.69	-1.20	2.40	-3.30	2.78	10.37
Sentence-initial coordinators	0.20	0.10	0.22	0.22	0.29	0.21	0.19	0.01	0.10	-0.01	-0.01	-0.09	0.22
Contractions	0.56	0.52	1.90	0.59	1.88	1.09	1.38	0.53	0.57	-0.81	0.50	-0.79	3.2
Mean word length	4.83	4.82	4.71	4.91	4.46	4.75	0.45	-0.08	-0.07	0.04	-0.16	0.29	0.64
Modality	1.06	0.73	1.10	1.27	1.14	1.06	0.54	0.00	0.33	-0.04	-0.21	-0.08	0.66
Imperative Mood	0.73	0.28	2.11	0.38	0.75	0.85	1.82	0.12	0.57	-1.26	0.47	0.10	2.51
Interrogative Mood	0.70	0.85	3.40	1.20	1.56	1.54	2.70	0.84	0.69	-1.86	0.35	-0.02	3.75
Pers. Pronouns (1st & 2nd person)	0.63	0.96	2.26	0.96	2.12	1.39	1.63	0.76	0.43	-0.87	0.43	-0.73	3.22
Nominalisations	2.30	2.20	1.62	2.73	1.08	1.99	1.65	-0.31	-0.21	0.36	-0.75	0.90	2.54
Passive Voice	7.33	14.10	7.52	8.52	6.38	8.77	7.72	1.44	-5.33	1.25	0.25	2.39	10.67
Boosters / Minimisers	0.80	0.56	0.97	0.68	0.95	0.79	0.41	-0.01	0.23	-0.18	0.11	-0.16	0.69
Titles	0.19	0.49	0.20	1.16	0.07	0.42	1.09	0.23	-0.06	0.22	-0.74	0.35	1.6
Eval. Adverbs / Adjectives	0.61	0.34	0.65	0.46	0.91	0.59	0.57	-0.02	0.25	-0.06	0.13	-0.32	0.78
Article Length	532.74	533.11	735.19	556.7	610.64	593.68	202.45	60.94	60.57	-141.51	36.97	-16.97	316.96