Y Pethau Bychain:
A functional description of the experiential structure of Welsh Nominal Groups

Lowri Williams
C0505911

Supervised by
Dr. Lise Fontaine

This dissertation is submitted in partial fulfilment of the requirements for the Degree of Master of Arts in Language & Linguistics

School of English Communication and Philosophy
16 January 2019

Word count: 16,433

Abstract

Berry (1975, p. 86) states that in order to answer some of the fundamental questions about language, it is first necessary to understand that language is a 'patterned phenomenon'. This essay will aim to describe the patterns and structures of experiential functions in Welsh Nominal Groups (NGs).

It will first outline the argument for the use of Halliday's Systemic Functional Grammar (SFG) as a theory through which it is possible to observe patterns of language in use, contrasting this with more formal approaches that have dominated Welsh syntax analysis to date. It will then outline the experiential features of NGs in English as outlined by Halliday and Matthiessen (2014), as well as demonstrating how descriptions of these phenomena have lead to advances in our understanding of language in both theoretical and practical contexts.

Next, it will respond to calls from Gibson and Fedorenko (2013) and Lundquist (2009), amongst others, for utilising empirical research methods in producing language descriptions. It thus builds on methodological approaches developed in the field of corpus linguistics to design a study in which the data is systematically collected, annotated and analysed. It will justify the use of *Cronfa Electroneg o'r Gymraeg* (The Electronic Corpus of Welsh) as a data source and explain how features were identified and operationalised.

In total, 766 Nominal Groups were analysed for functional elements and structures in Subject position. This revealed that Welsh NGs contain all six of the functional elements identified by Halliday (Thing, Deictic, Numerative, Classifier, Epithet and Qualifier). Their position and distribution within the NG suggest that, in Welsh, the least specific function sits at the core of the NG (i.e. the Thing) with functions becoming more specific and less 'permanent' as they radiate out from the centre of the NG.

It concludes by exploring the consequences of this analysis both for our understanding of the language, but also in contributing to the SFG theory as a whole.

Acknowledgements

Firstly, I would like to express sincere gratitude towards my supervisor Dr. Lise Fontaine for the encouragement, wisdom, and kindness she has shown me throughout this MA year and in particular in the preparation of this dissertation. I was first introduced to Lise's work through the Open University several years ago, and it's been a privilege to be able to come to Cardiff and work with her in person. I couldn't have hoped for a more understanding (and patient!) mentor.

Secondly, I'm eternally grateful for the support of my partner, David, throughout this year and the years that have led up to me completing my Master's. Thanks also for his practical help in proof reading final drafts of this dissertation and for supporting me through technical blips. I'm sorry I made you use Windows.

Thanks to my fellow students – the self-proclaimed *Sweaty Linguists* – for their friendship and encouragement throughout the course. Also, to my colleagues at *Corpws Cenedlaethol Cymraeg Cyfoes* (CorCenCC) and to Gwen Rice at *Welsh for Adults* who, along with this dissertation, have reignited in me my love of the Welsh language.

Finally, thanks to Cardiff University for operating the Master's Excellence scheme, making it possible for students like me to return to full-time education.

Diolch o galon.

Contents

Declarations ..... **Error! Bookmark not defined.**
Abstract 3
Acknowledgements 4
Abbreviations 7
General Terms 7
Part-of Speech Tags used by Cronfa Electroneg o'r Gymreag 7
Chapter 1 Introduction 8
Chapter 2 Literature Review 12
2.1 Introduction 12
2.2 Form versus function: the case for SFG 13
2.3 Referring expressions 16
2.4 Nominal Groups in English 18
2.5 The Welsh Language 21
2.6 'Standard' Welsh 22
2.7 Studies in Welsh grammar to date 23
2.8 Nominal Groups in Welsh 25
2.9 Conclusion 25
Chapter 3 Research Methods 28
3.1 Introduction 28
3.2 Research Strategy 29
3.3 Research Design 33
3.3.1 Data collection 33
3.3.2 Determining a sample 34
3.3.3 Building and manually tagging the data sample 35
3.4 Analysing the data 36
3.4.1 Identifying grammatical Subjects 36
3.4.2 Identifying NGs 37
3.5 Limitations 39
3.6 Summary 40
Chapter 4 Functional description of NGs in Welsh: corpus findings, analysis and discussion 41
4.1 Introduction 41
4.2 Realization of grammatical Subject 42
4.3 Thing 47

4.3.1 Things realized by nouns 47
4.3.2 Things realized by other word classes 49
4.4 Deictic 51
4.5 Numerative 57
4.6 Epithets 61
4.7 Classifiers 63
4.8 Qualifiers 63
4.9 Ordering of functional elements in the NG 65
4.10 Conclusion 65
Chapter 5 Conclusion 68
5.1 Introduction 68
5.2 Potential areas for future research 71
5.3 Contribution to knowledge 72
References 74
Appendices 77

Abbreviations

General Terms

CEG	Cronfa Electroneg o'r Gymraeg – Corpus of Digital Welsh
NG	Nominal group
NP	Noun phrase
POS	Part-of-Speech
SFG	Systemic Functional Grammar
SFL	Systemic Functional Linguistics

Part-of Speech Tags used by Cronfa Electroneg o'r Gymreag

nm	singular masculine noun
nf	singular feminine noun
nm/f	singular noun (masculine or feminine depending on context)
npl	plural noun
pron	pronoun
vb	verb
card	cardinal number
DemPron	demonstrative pronoun
adj	adjective
prep	preposition

Chapter 1 Introduction

"Gwnewch **y pethau bychain** mewn bywyd**"**

("Do **the little things** in life")

Traditional Welsh saying (emphasis mine)

When these words were reportedly uttered by St. David, the patron saint of Wales, on his deathbed in the 6th century, it is unlikely that grammatical analysis was at the forefront of his mind. Yet this principle is certainly applicable today when approaching the great 'puzzle' of language: if we wish to tackle some of the bigger questions, it is sometimes necessary to start by understanding the smaller details. The focus of this thesis is precisely that: investigating both figuratively and literally the 'little things' that shape, and are shaped by, our experience of the world around us.

The 'little things'[1] in question are Nominal Groups (NGs)[2] in Welsh. Fontaine (2012b, p.44) defines NGs as types of 'referring expressions': grammatical structures that allow us to conjure in the minds of our audience any 'entity' that we want to say something about. Entities can be human or abstract, objects you can hold or concepts of the mind. In

[1] The term 'thing' is used here in its general, everyday sense as indicated by the lowercase initial. This is to distinguish it from the more specific meaning of Thing as a functional element of an NG in SFG (explained in Chapter 2).

[2] 'Nominal Group' is the term most commonly used across SFG (e.g. Fawcett 2000, p. 203 - 204; Fontaine 2012b, p. 46; and Halliday and Matthiessen 2014, p. 364 - 395). Other theories, including the Transformagtional-Generative approach (TG), instead use the term 'Noun Phrase' (e.g. Borsley et al. 2007a). As SFG is the theory adopted in this thesis, 'Nominal Group', or NG, will be used throughout unless specifically stated.

analysing the ways in which NGs are structured, it is possible to understand the choices available to a speaker of a language when they wish to refer to something in particular.

One means of achieving this is through Halliday's Systemic Functional Grammar (SFG) – a theory which puts meaning, function and context at the centre of grammatical analysis (Fontaine 2012a). For SFG, the language we speak and the situations in which we communicate all shape the choices available to us when referring to particular entities. While the NG remains a little-explored corner of SFG relative to some of its other concerns, functional descriptions of NGs in English exist (cf. Fawcett 2000; Halliday and Matthiessen 2014; Fontaine 2012b), providing well-needed insights into how we talk about the things in our world. Furthermore, such insights have led to practical applications of NG research, such as in foreign language teaching (e.g. Moalla 2018).

Fontaine (2007), however, stresses that there is no requirement for all languages to be patterned in the same way. Indeed, even a cursory glance at French or Chinese would confirm this to be the case. There is a need, therefore, to be inclusive of as many varieties of language as possible if the theory is to be robust and flexible (Caffarel 2006).

Proposing a functional description of the structure of Welsh NGs is therefore motivated by three primary factors: firstly, to incorporate into the SFG a minority language that has yet to have undergone thorough examination of one of its features through a functional framework. While descriptions of NGs in Welsh are available, they has thus far been limited to one sphere of syntax study, that of the Transformational-Generative approach (TG) (i.e. Borsley et al. 2007). As will be discussed in Chapter 2, this is just one approach with a

defined set of objectives, limiting the questions that might be asked of the data (Thompson). Secondly, in shifting to a functional perspective, it is possible – and appropriate – to respond to calls from Lundquist (2009, p. 8), Sampson (2001) and Gibson and Fedorenko (2013) for more a more systematic, transparent and verifiable dissemination and analysis of data than has traditionally been undertaken by TG approaches of the past. Responding so these concerns, this study will draw on experiences and recommendations from the field of corpus linguistics (e.g. Hunston 2013; Neale, O'Donnell 2014) to produce a description of Welsh NGs to meet this criteria.

The thesis is structured as follows:

Chapter 2 Literature Review will introduce the concept of grammar theory, establishing the motivation for observing 'patterned phenomena' of language through the lens of syntax (Berry 1975, p.86). It will discuss the 'form versus function' debate, contrasting the approaches of Chomsky and Halliday before justifying SFG as the framework of choice. It will then provide an overview of the internal structures of the NG in English according to Halliday and Matthiessen (2014, p. 364), explaining how the individual structures relate to both function and form. Finally, it will provide an overview of the Welsh language, TG descriptions of the NG completed to date, as well as an insight into the challenges posed in attempting to describe a language with limited resources. It will conclude by introducing the key research aims, namely the production of a functional description of Welsh NGs.

Chapter 3 Research Methods will outline the detail the methodological approach to data collection taken by this study. It will do this by firstly evaluating the argument for more empirical approaches to data collection, dissemination and analysis such as those posed by Gibson and Fedorenko (2013) and Lundquist (2009). It will then outline the argument for

incorporating corpus-based methods into SFG study, citing developments technology that can help overcome some of the complexities of SFG annotation. Next it will discuss the research design, from data collection, dissemination to analysis.

Chapter 4 Functional description of Welsh NGs: corpus findings, analysis and description will then present and evaluate the main findings of this joint SFG and corpus-driven investigation. The key objective is to provide a functional description of elements present in the Welsh NG. It will begin by reporting on the quantitative observations in the form of frequencies of functional patterns identified in the text, along with the formal word classes that realize these choices. These will be supplemented with qualitative data in the form of examples taken from the dataset. The chapter will detail each element in turn, drawing on descriptions proposed for English NGs by scholars such as (Halliday and Matthiessen 2014) and (Fontaine 2012b). An analysis of these findings will follow, exploring the consequences of such patterns for how meanings are conveyed in Welsh. Finally, it will propose a description of NGs in Welsh based on these findings.

Chapter 5: Conclusion will close this study by briefly summarising its key findings in the context of the research questions posed, and evaluating what they mean for the Welsh language and the SFG framework. It will evaluate the value of empirical data methods as a means of producing a functional description. Recognising the limitations posed by the study, it will propose a series of next steps both in terms of theoretical work and practical applications. It will then conclude by reflecting on the value of including minority languages in SFG analysis, and the benefits of analysing such data using empirical methods.

Chapter 2 Literature Review

2.1 Introduction

According to Berry (1975, p. 86), in conducting linguistic research one is generally concerned with answering some fundamental questions about language: '*What is language? How does language work? How are languages different?*' Attempting to answer any or all of these questions requires recognition of the fact that language is a 'patterned phenomenon' and that the ways in which patterns are produced and repeated can reveal features of the language that might otherwise have gone undetected (Berry 1975; Baker 2010, p. 94). This chapter will put forward the case for identifying some of the 'patterned phenomena' of the Welsh language, with particular focus on 'referring expressions' or Nominal Groups (NGs) (Fontaine 2012b, p. 44).

It will begin a review of the literature by introducing the concept of syntax theory as a means of interrogating linguistic patterns. It will discuss two central frameworks in the field, that of Chomsky's Transformational-Generative (TG) approach (e.g. Chomsky 1965), contrasted with Halliday's Systemic Functional approach to Grammar (SFG) (e.g. Halliday and Matthiessen 2014). This will ultimately outline the argument for approaching syntactical analysis 'through function', as opposed to observing patterns in word classes alone (Thompson 2014, pp. 2-11).

It will then review Halliday and Matthiessen's (2014, p. 364) description of NGs in English, drawing on supporting descriptions from Fontaine (2012b) and Bloor and Bloor (2004). It will introduce the functional elements that Halliday predicted would be present in NGs - namely the Thing, Numerative, Deictic, Classifier, Epithet and Qualifier – along with how these functions interact to allow a speaker to refer to a specific entity (Halliday and

Matthiessen 2014). It will also explore NG analysis carried out in other languages, as well as providing an overview of some of the practical applications of NG analysis, such as in foreign language teaching (e.g. Moalla 2018).

Finally, it will evaluate syntactical work carried out to date on NGs in Welsh incorporating academic research from other theories (i.e. Borsley et al. 2007a) as well as popular reference grammars such as King's *Modern Welsh* (2003). In doing so, it will also highlight some of the practical and theoretical obstacles that face a potential researcher in this field. Ultimately, it will propose that the inclusion of Welsh in the SFG framework would be beneficial to both the language and the theory, and that the Nominal Group presents itself as the ideal starting point for such a description. It will close by detailing two research questions to frame this description and analysis.

2.2 Form versus function: the case for SFG

At this juncture, it would be sensible to provide some theoretical context to support the use of grammar as a framework through which to understand language. Syntax, the study of grammatical rules and principles within a language, developed significantly in the 20th century, producing branches and frameworks from which the exercise of mapping out and describing grammatical patterns could be undertaken. One of the more dominant approaches has been the study of 'traditional' grammar classes, particularly that of the Transformational-Generative (TG) approach championed by Chomsky (1965). TG grammarians describe grammatical rules of a language in terms of the patterning of 'formal' components, i.e. nouns, adjectives and their wider groupings such as noun and verb phrases. A key objective of TG is to describe the language of an 'idealised speaker', one that is unencumbered by context and the inherent fallibility of being human (Lundquist 2009,

p.8). As such, it is very much a theory that aims to describe linguistics in terms of accuracy and accessibility. Crucially, Chomsky believed that context was an unnecessary distraction, arguing that speakers frequently express language imperfectly therefore observing 'real' language would be fruitless. Instead, traditional syntacticians tend to produce grammatical descriptions of a language via introspection, that is to draw upon their own knowledge of language and what can be gleaned from resources about grammar without explicitly calling upon observed data (Lundquist 2009).

Halliday's approach - Systemic Functional Grammar (SFG) - takes issue with some of the key principles of TG, in particular its dismissal of context as a key role in language production. On the contrary, Halliday argued that language both construes and is construed by the context in which it is being produced (Halliday and Matthiessen 2014, p. 4). This is to say that a producer of language – henceforth the 'speaker' – is continually making choices about their words and grammar in response to the environment in which it is taking place. Halliday refers to this as 'lexicogrammar', which can be recorded, observed and analysed for grammatical and functional patterns (Halliday and Matthiessen 2014, p. 7). A speaker giving a political speech, for example, will likely make different choices about their language than they would addressing their young children at the dinner table. Likewise, language can be said to actively shape the environment in which it occurs. Rather than being products of our human imperfections, Halliday argues that context is an important part of our investigation, and to do that we must aim to analyse real, observable examples (Halliday and Matthiessen 2014, p. 29).

Another distinguishing feature of SFG is its focus on 'function' over 'form' (Thompson 2014, p. 2-11). Thompson argues that 'going in through form', as is the TG approach, limits the

types of questions one can ask about language. While observing patterns in word types can reveal its underlying rules, it is lacking in that it does not explain why such rules exist, and what function is being fulfilled by each component. To this end, an alternative approach is needed to categorise and interpret grammatical patterns according to the function they undertake in the text. SFG is one means in which this is achieved.

Systemic Functional Linguistics (SFL) is the theory to which SFG belongs. It is considered a 'systemic' theory in that it proposes a model for language in which the language we produce is considered the output of a chain of ever more delicate decisions (Halliday and Matthiessen 2014, p. 24). Choices may be limited or shaped by specific conditions, such as a formal or informal context, or the purpose of the text. These decisions therefore appear in a system of complex decisions, in which selecting one options necessarily excludes others. When one considers this process as decisions become more and more specific or 'delicate', the words that are produced as a result are the output and not the whole story. This is why, as Thompson (2014, p. 4) posited, a researcher would be limited in the questions they could ask of their data, were they to 'go in through form' alone. By extension, SFG is also a paradigmatic theory. This means that the output can be seen as boxes that need to be filled, where the ordering of the boxes is predetermined, but that leaves choice in what word is chosen within it depending on their motivations and the situation in which they speak.

Finally, SFG is 'functional' because, as previously discussed, the theory is particularly concerned with language in use (Halliday and Matthiessen 2014, p. 64). The researcher's role is in trying to understand why and how language behaves in this way and how the context of the situation and the speakers within it may have influenced the decisions made. Halliday argues that language at any given time is fulfilling three overarching functions - or

'metafunctions': the first is the ability to describe something that is happening (the experiential metafunction), the second describes the relationship between the speaker and audience (the interpersonal metafunction) and finally the Textual metadunction concerns the role of the language in forming the text (Thompson 2014, p. 31). These metafunctions are all tightly interconnected, each occurring simultaneously but often approached independently for analysis. Ultimately, the researcher is concerned with what the language is doing in this context, over and above the actual choice of words used.

Language, to Halliday, is therefore a 'meaning making resource': allowing us to build relationships, create meaningful and understandable texts and, crucially for this thesis, share our experience of the world with others (Halliday and Matthiessen 2014, p. 86). SFG is concerned with language in use, preferring to observe patterns in terms of probability instead of making judgements about acceptability (Berry 1975). It therefore calls for a systematic approach to data analysis which has been absent from other approaches. As Berry suggests, understanding some of the key issues concerning language is the ability to draw out the patterns that occur within such texts, and evaluating them in terms of the context in which they appear, the structures in which they occur and how these resources relate to the wider language puzzle. One way of doing this is understanding how a language allows us to talk about 'things'.

2.3 Referring expressions

Things dominate our language. According to research, they are the types of linguistic expressions first picked up by bilingual children and the type of word most frequently put into search engines (Chan and Nicoladis 2010; Jones et al. 2016). The ability to refer to an entity is therefore a vital part of how we communicate our experience of the world with

others, i.e. Halliday's 'experiential' metafunction. In language, this is achieved through 'referring expressions':

*"...we can define a referring expression as a **linguistic expression** that the speaker uses to **refer to an entity which is a participant in the situation** he or she is describing."*

(Fontaine 2012, p. 45, emphasis mine)

In terms of the entities to which a speaker refers, it helps to refer to the concept of Subject in more traditional grammar (Fontaine 2012). When evaluating a segment of speech – a clause - grammatical Subjects generally answer the question 'Who or what is doing the action?'. If we follow that every clause must have a Subject, and that Subject is at the core of the activity taking place, it is possible to appreciate the importance of referring expressions in understanding not only the language itself but how its speakers characterize their experiences of the world.

Traditional schooling would tell us that we refer to entities through nouns – what Bloor and Bloor (2004, p. 18) refer to as 'naming words'. However, as previously discussed, SFG sees words as the final, most delicate of the systemic decisions, therefore cannot hope to describe the full picture. It is therefore the entire cluster of components that hold the function of specifying the item we're referring to, rather than individual word classes. In English, this function is most commonly realised via the Nominal Group (NG) (Fontaine 2012b, p. 46).

For Halliday and Matthiessen (2014), NGs are 'a meaning-loaded grammatical unit': they allow us to explore the three 'metafunctions' central to the theory as it is commonly upon the Subject that all three functions conflate. The experiential structure of the NG and the

ordering of its components allows the speaker to signal an entity they have in mind, making choices from the linguistic and systemic options available to them at that juncture to ensure that this entity can be correctly identified and recreated in the mind of their audience (Halliday and Matthiessen 2014, p. 654). Understanding and describing this structure is therefore a key step towards understanding a language as a whole.

2.4 Nominal Groups in English

Perhaps unsurprisingly, it is English that has thus far received the most attention in terms of NG analysis in SFG (e.g. Halliday and Matthiessen 2014; Fontaine 2012b; Fawcett 2000). It has been posited that one can approach this analysis from two different perspectives: the: intra-organism orientation – what Halliday calls "language from the outside"; and inter-organism orientation - language from the "inside" (Fontaine [forthcoming], p. 2). The former refers to the way that NGs interact with each other and with other groups in language, linking theoretical study to concerns around society and discourse. The latter looks more at the mechanics of the elements itself, exploring what each individual feature is there to achieve and why it has been structured in such a way. It is here that patterns of experiential functional patterns might be best explored.

As previously suggested, it is not a word class alone, e.g. noun, that carries the function of referring, it is the NG as a whole. These may be realized in formal terms by nouns or adjectives and so on, but these are simply outputs. At the core of an NG is the specific class of item, concept or entity that a speaker is trying to conjure in the mind of their audience. Depending on need, these might be accompanied by other items, to further describe or specify that core entity. According to Halliday, these are the functions present within the NG in English:

1) **Thing:** This is the 'semantic core' of the NG, representing the class of an entity in its most general sense (Halliday and Matthiessen 2014, p. 364). It is said to be the only required element of the NG in English, meaning that a Thing might be realized simply with a noun, such as in 'train' or 'thoughts' (Halliday and Matthiessen 2014, p, 383; Fawcett 2000). In terms of form, Things are commonly represented as nouns. Where a Thing requires further specification, to identify a particular subset class of train, e.g. 'steam train' or 'my train', Things are accompanied in the NG by other functional items.

2) **Deictic**: This carries the 'pointing out function', that is that it signals whether or not a 'specific subset of the Thing is intended' (Halliday and Matthiessen 2014, p. 366). This function is realised by determiners such as 'the' or 'my' which indicate the specificity of the Thing in question. Specific determiners are split into demonstrative determiners realized by definite or indefinite articles (e.g. 'the' or 'an') or possessives realized by personal pronouns (e.g. 'my'). These can also denote proximity to the Thing in question, such as 'these' or 'those': allowing the speaker to refer to context or entities already accessible in the text.

3) **Numerative**: According to Halliday (ibid., p. 374), the Numerative 'indicates some numerical feature of the particular subset of the Thing', allowing the speaker to express order (e.g. first), amount (e.g. five) and quantity (e.g. every). These are typically realized by numerals in formal grammar.

4) **Epithet**: These function to describe a quality of the Thing, making it identifiable through its descriptive characteristics. These are commonly realised by adjectives, and have the ability to be intensified, e.g. very big, differentiating it from the other type of modifying function – the Classifier.

5) **Classifier**: In English, the Classifier function allows Things to be classified in some way, referring to a sub-class of the Thing in question. For example, a 'bus station' specifically identifies a station for buses as opposed to a 'train station' or 'police station'. These can be realized by both nouns and adjectives but cannot be intensified in the same way as Epithets e.g. '*a very bus station' (Halliday and Matthiessen 2014, p. 378). Finally, in terms of order, Classifiers follow Epithets when both are present, for example 'busy bus station'.

6) **Qualifiers:** These appear at the very end position in the NG, which in English is directly after the Thing. They function to characterize the Thing in some way, putting it into a specific context that specifies it from more general instances of the Thing. Qualifiers differ from the other functional groups as they are most frequently realized through embedded phrases or clauses, and not individual words. In English this is most frequently through a prepositional phrase (e.g. 'the girl with the big hair'), an embedded relative clause (e.g. the house that Jack built) and, less frequently, an adjective (e.g. 'president elect').

Example [1], from Halliday and Matthiessen 2014, p. 364), demonstrates a possible NG structure in English, along with the formal word classes that realize these functions.

[1]

those	two	splendid	old	electric	trains
Deictic	Numerative	$Epithet_1$	$Epithet_2$	Classifier	Thing
determiner	numeral	adjective	adjective	adjective	noun

This example also illustrates Halliday's argument that the way in this these functions are structured tells us something about the way refer to entities in English. He argues that "there is a progression in the nominal group from the kind of element that has the greatest

specifying potential to that which has the least" (Halliday and Matthiessen 2014, p. 381). This is to say that the NG gets less specific the further right we move through the NG towards the Thing.

In observing functional patters, this links to the idea that SFL, and by extension SFG, is paradigmatic. This is to say that while there is an underlying functional structure to NGs, and that order cannot be altered in most cases, how they are realized leaves room for the creativity we need to be able to express the multitude of entities we discuss every day. It is here that frequencies of word classes might be analysed, to determine the probability that a certain word class (such as a noun) would be selected to realize the Thing. It also allows us to minimalise ambiguity, a key function of interpersonal language, because we understand the underlying functions that are being expressed.

Fontaine (2007)reinforces the importance of NGs in language description, suggesting that all languages need the ability to refer to 'things' There is no reason to expect, however, that this would be achieved in the same way as it is in English: 'the structural realization of a referring expression is not a given' (ibid., p. 161). Analysis of NGs in other languages such as Welsh is therefore a worthwhile pursuit, as it may reveal functions and features of language previously unreported, challenging existing theories.

2.5 The Welsh Language

Welsh is understood to be the oldest living language in Great Britain and Eire (King 2003, p.1). A member of the Proto-European language family, it is the most widely spoken of the surviving Celtic languages with 28.1% of the Welsh population stating they can speak some level of the language (Welsh Government 2018). Its closest linguistic relatives are said to be the Brythonic Celtic languages found in Brittany (Breton), Cornwall (Kernowek) and the Isle

of Man (Manx). Though more distant, there are also commonalities between the Goedelic languages of Irish and Scottish Gaelic. Lexically, the Celtic languages share cognates as well as bearing hallmarks of earlier Roman influence such as *pont* ('bridge') and *ffenestr* ('window'). Grammatically, the Celtic languages are linked, amongst other things, by their initial mutation patterns: where the first letter of affected words will change to synthesise with the word that precedes it. They are also said to grammatically be of Verb-Subject-Order, with the verb, auxiliary or verb phrases in initial position in the clause (see **Error! Reference source not found.** adapted from Borsley et al. 2007b, p. 341). Such characteristics therefore pose a challenge for existing theories as they challenge dominant narratives based on English.

[2]
Gwelodd y ddynes ddraig
see.PAST.3S the woman dragon
'The woman saw a dragon

2.6 'Standard' Welsh

The diverse linguistic landscape of modern-day Wales was carved by its physical landscape, along with its history of invasion and industry. The result has been a language with a wide range of dialects and variation. Differences between North and South Welsh, for example, are widely recognised in Welsh language learning. These differences are believed to be mainly lexical, such as the translation of the word 'money' in which North Welsh speakers use *pres* ('brass') while their South Walian counterparts pay with *arian* ('silver'). King (King 2003, pp. 2-5), however, argues that these differences are not as clean cut as North and South, and there are even some subtle grammatical differences between the two.

Attempts have been made throughout its history to create a standardised version, most notably resulting Literary Welsh which was established with the translation of the Bible around (King 2003). Due to the religious and officious context in which it was being used, it naturally excludes many of the rich features spoken amongst working class Welsh speakers as well as dialectical variance. It has also been criticised for devising new and complex rules that had not existed in the spoken language until this point, rendering it too unwieldy for spoken discourse. As a result, present day Literary Welsh is precisely thus – almost exclusively banished to formal literary contexts such as those encountered at Eisteddfodau, traditional cultural celebrations. Efforts to standardise the language were revised in the 1970s, resulting in what has been largely called 'Cymraeg Byw' (*Living Welsh*) (King 2013). Broadly speaking, this is the variety of language seen in news reports and formal, non-literary writing. Spoken Welsh, however, still varies greatly, it is widely acknowledged that no standard version of language has yet sufficiently handled these regional and contextual nuances. This presents the syntactician with the daunting task of describing a language in which there is no recognised 'standard'.

2.7 Studies in Welsh grammar to date

Extensive literature searches revealed that little has been published about Welsh through Halliday's Systemic Functional Linguistics framework. In their meta-study of published SFL work, Mwinlaaru and Xuan (2016) identified just one study incorporating Welsh phonology. To date, there appears to be no published studies that have explored its grammar functions specifically.

To understand Welsh syntax, it has therefore been necessary to expand the search for beyond the boundaries of functional grammar. What has been done on Welsh syntax falls

largely into two categories: reference grammars and generative grammar. Reference grammars act much like a dictionary, a 'go to' reference for enquiries about grammatical rules and patterns. Their audience is an interested but not necessarily academic reader, and are thus published as books for the public, not requiring the levels of peer review or methodological scrutiny that might befit an academic publication. These are by far the most abundant and accessible source of information on Welsh grammar. Several versions exist, but certainly the most well-known and widely available is King's *Modern Welsh* (2013). King makes no claim to adhering to any grammar theory, and focusses only on form, not any meaning that might be derived from it. It is, however, a comprehensive and systematic description of Welsh, making it a useful reference to accompany investigations into the field.

The second group of resources are the syntactical descriptions published in academic journals and books. These vary in specificity, ranging from more detailed problems such as querying whether the 'verbal noun' in Welsh is actually a verb or a noun (e.g. (Willis 1988; Hannahs and Tallerman 2006; Borsley et al. 2007a). Importantly, these materials are written entirely through the framework of generative linguistics, whether or not this fact is explicitly or implicitly declared. As established earlier in this review, Transformational-Generative (TG) approaches, and related generative frameworks, are necessarily focussed on the grammar forms: nouns, adjectives, and so on. There is therefore no compulsion to explore the function of language patterns, nor is there a requirement for observable data. While the consequences of this will be discussed in depth in **Chapter 3 Research methods**, it is nonetheless a necessary to bear this lack of transparency and data integrity in mind when drawing upon such resources to build a description of a language.

2.8 Nominal Groups in Welsh

At time of writing, there exists one comprehensive analysis of the structure of Welsh NGs[3] is via the Transformation-Generative approach adopted by Borsley et al. (2007a) in their book *The Syntax of Welsh*. The authors claim that the NG follows the structure Determiner, Numeral, Noun, Adjective, Possessor/Demonstrative, Complement and Relative Clause (ibid. p. 152). As expected of TG approaches, this structure is described according to formal features, and lack the functional insight provided by SFG.

2.9 Conclusion

This chapter has presented the motivations for commencing the analysis of Welsh syntax by producing a functional description of the functional elements of its NGs. We began by exploring the wider theoretical landscape concerning the study of grammatical patterns, introducing the broadly contrasting theories of Chomsky and Halliday and their stance on the role of context and data in producing linguistic descriptions. It went on to discuss the debate around 'form over function' and why, as Thompson suggested, the use of one framework over another is not a question of being 'better', rather should be seen as a template for the kind of questions one might ask of our texts in a bid to demystify the puzzle of language.

It went on to identify Halliday's Systemic Functional Linguistics, and SFG more specifically, as an appropriate framework for posing questions around function, structure and meaning, insights that have yet to be applied to Welsh language data thus far. By extension, it is therefore impossible, in Halliday's view, to divorce language from its context, because a speaker is necessarily making up-to-the-minute decisions about lexicogrammatical choices

[3] Note that, as previously mentioned, the term used in TG analysis is generally 'noun phrase' or NP

in response to the environmental situation in which it is taking place. It is therefore Halliday's view that real data, such as that collected via a corpus, that would help observe linguistic patterns in context and build a system in which choices are described according to their probability of occurring, rather than their acceptability.

Next, the Nominal Group (NG) was singled out as a reasonable starting point for analysis into a hitherto 'unmapped' language for SFG. This was arrived at according to their abundance in language, most commonly as participants in whatever action is taking place within the clause. An overview of NG analysis carried out to date revealed the dominance of English in this field, with much of Halliday's seminal works on the feature done in respect of English. The English NG's individual features were identified and described, explaining how these relate to the experiential function of the NG. Furthermore, it was shown that understanding NG structure and function had practical applications in fields such as language learning, equipping teachers with additional tools and perspectives to understand and evaluate the progress of their students. There exists, therefore, a need to expand this understanding of NGs to many more languages, not only to improve our understanding of the language itself, but to force the theory to become as robust, inclusive and adaptable as possible.

The latter part of this literature review was concerned with introducing the Welsh language and work done to describe it thus far. It revealed a number of issues for the potential syntactician, namely the lack of a standardized vernacular, as well as a dearth of academic resources dedicated to its grammar. It briefly described work carried out through the TG route, namely that of Borsley et al. (2007a).

In summary, this literature review has established the need for Welsh to be analysed according to the SFG framework, and that NGs are a sensible place to start given their important role in construing experience through language. Crucially, there is a need to address this in a way that can be empirically verified, avoiding some of the issues of validation and repeatability that might befall introspective descriptions carried out to date. I therefore propose the following research questions for exploration in the following chapters:

- What are the functional elements of Nominal Groups in Welsh?
- What is the experiential structure of Welsh NGs, and what does this say about the way Welsh refers to entities?

Chapter 3 Research Methods

3.1 Introduction

As outlined in the Literature Review, the primary objective of this study is to propose a theoretical description of the functional structure of Welsh Nominal Groups (NGs). In doing so, it will answer the two research questions posed in Chapter 2:

- What are the functional elements of Nominal Groups in Welsh?
- What is the experiential structure of Welsh NGs, and what does this say about the way Welsh refers to entities?

A survey of the literature revealed three key issues with the production of descriptions of Welsh to date: firstly, there are currently no published descriptions of Welsh grammar conducted via Halliday's Systemic Functional Grammar (SFG) approach. Secondly, that the literature is limited to reference grammars and descriptions written through the Transformation-Generative approach championed by Chomsky. The implication being that both approaches generate descriptions based on introspection alone: drawing on the author's own knowledge of the language to document its rules in terms of accuracy, rather than capturing language in use. Finally, that the Welsh language poses problems for any researcher attempting to capture a 'standard' version, as attempts to do so have resulted in the exclusion of the spoken and regional variety characteristic of the language and its speakers.

This chapter will therefore propose a methodology designed such that it will respond to and mitigate these three concerns. The result will be functional description derived through a

transparent data collection, dissemination and analysis process. Section 3.2 will discuss the motivation behind adopting an empirically-focused data analysis approach. It will evaluate criticisms of the introspective approaches deployed to date, responding to calls from Gibson and Fedorenko (2013), Lundquist (2009) and Baker (2010), amongst others, to hold syntax study to the same scientific standards of other social science fields. Section 3.3 will then document the research design. It will begin by discussing the approach to data collection, before explaining how the data was operationalised and disseminated to enable functional analysis to take place. Finally, it will detail some of the limitations of a small-scale study, outlining ways that the effects of these limitations will be mitigated.

3.2 Research Strategy

Lundquist (2009) argues that linguists should hold themselves to the same rigorous standards expected of other social science disciplines. They are especially critical of introspective methods deployed in linguistic theory to date: approaches which enable the researcher to draw on their own understanding of language to produce descriptions. The risk, they argue, is that in doing so they run the possibility of excluding or misrepresenting patterns in language, even creating convenient examples to prove a given point. Our aim should be an inductive approach, to build descriptions based on observations in the data. They therefore urge linguists to move away from the comfort of using introspective methods alone because by embracing a methodology that uses real data, linguistic phenomena can be extrapolated systematically to avoid researcher bias. This should result in robust studies that can be repeated and verified, meeting the demands of scientific enquiry. This is a view shared by Gibson and Fedorenko (2013). They argue that, within syntactic theory in particular, there has not been sufficient development on this front.

Indeed, they reference instances in which the introspective approach has resulted in inaccurate conclusions (ibid., p. 233).

Corpus linguistics (CL) has been proposed as one possible means of approaching such a task (Hunston 2013). A corpus is a body of texts that can be said to be representative of language in some way (Baker 2010, p. 93). This might be the language as a whole (a 'general' corpus) or one dedicated to a particular aspect of it, such as genre (Baker 2010). Corpora can be annotated and analysed manually, but in the modern age researchers commonly deploy the use of dedicated technologies. Resources such as AntConc and Sketch Engine are widely available to assist with the organising and extrapolation of phenomena, but equally this can be done via Excel. Baker, citing McEnery and Wilson (1996, p. 1, in Baker 2010, p. 93), explains that corpus linguistics is a methodology rather than a linguistic theory in its own right. This means that it is possible to use CL as a tool through which we can explore, challenge and evidence linguistic theories such as SFG. Furthermore, Stubbs (2006, p. 24) defines CL as 'an empirical approach to language study' as it allows the researcher to observe patterns in volume, using real or 'authentic' data. This could satisfy Lundquist's plea for more inductive research approaches, in doing so raising 'new questions and theories about language that otherwise would not have been possible' (Baker 2010, p. 94).

CL thus affords the researcher options to interrogate data quantitatively and qualitatively, depending on the questions that might be asked. In operationalising our data, grouping or 'coding' observations into categories based on shared characteristics, it is then possible to begin observing patterns of frequency (Levon 2010, p. 69). This might be frequencies of words, frequencies of word combinations or frequencies of word classes such as parts of speech (POS) (Baker 2010, p. 97). It also provides the option to qualitatively observe

features in their immediate textual context, or co-text, via the use of concordance lines (Stubbs 2006, p. 24). Studies are often characterized as being corpus-based or corpus driven depending on whether one is looking to the corpus to build a hypothesis or to test one. However, as (Angouri 2010, p. 33) points out, there is no need to restrict ourselves to one theory or one approach. Indeed, combining perspectives and methodologies can result in a 'diversity of views', which can only strengthen the inferences that we make. Using corpus data to produce an estimate of frequencies can be accompanied by closer, more qualitative approaches required by SFG to understand the context in which the lexicogrammar appears. Furthermore, McEnery points to the false dichotomy of 'corpus-driven' versus corpus'-based, stating that, in fact, a researcher most often falls somewhere in between.

Corpus linguistics in SFG

Halliday himself advocated for the use of corpora in SFL study, arguing that a lack of data has held back SFG study in the past (Halliday and Hasan 2000, p. 201). At the heart of the theory is the desire to describe linguistic phenomena in terms of frequencies of probability, taking instances of language in use as evidence of linguistic phenomena (Neale 2006, p.145). It is therefore unsurprising that focus in recent years has shifted to incorporate CL methods in SFG (Hunston 2013). Indeed, corpora have been used in functional grammar study since the 1950s, but this has become more accessible due to recent developments in corpus software (O'Donnell 2014, p. 346).

Due to the highly contextual nature of SFG theory, however, it has not yet been possible to automatically tag texts for functional features. This is because texts consist of words that are the output of functional descriptions, and it is therefore difficult to manually tag for these without understand the context in which it occurs. O'Donnell (2014) recognises this

limitation for deploying CL methods in SFG study. As Gibson and Fedorenko (2013) point out, incorporating data methods such as these can be time consuming and complex. Indeed, in Moalla's (2018) study into how Arabic speakers translate NGs into English, the use of authentic data was considered but later dismissed due to the complexity of real language in use. However, both O'Donnell and Gibson and Fedorekno argue that overcoming such obstacles is worthwhile if it produces systematically verifiable results. It might therefore be appropriate to restrict our focus to particular types of clause, or to opt for a smaller sub-corpus of texts to reduce the manual effort required. Sampson (2001, p. 2) and Baker (2010) are keen to point out that using a small dataset does not necessarily mean that a smaller corpus would result in less accurate results. Indeed, they sight cases in which successful and insightful analysis has been carried out on very small sets of data, such as Stubbs (1996).

Neale (2006) also recognises the limitations of SFG/CL method design, particularly the point that available corpora are rarely if ever coded for SFG functions. She therefore proposes a 'second-level of study' approach, using corpora tagged for other means, such as formal parts-of-speech (POS), to help shape and focus SFG study. This approach, adopted in their study of transitivity patterns, involves building on patterns observed in part-of-speech (POS) tagged corpora and then making decisions based on SFG. The researcher is then able to make inferences not only about functional grammatical patterns, but the probability of being realized in some way by formal grammar features. This reinforces Angouri (2010, p. 34)'s argument that triangulation – looking at data through different frameworks – can lead to a better understanding of the phenomena with which we are interested.

3.3 Research Design

Reflecting on the discussion in Section 1.2, this section will now outline a research approach that incorporates corpus-based methods into the description of Nominal Groups in Welsh.

3.3.1 Data collection

Selecting a corpus

Ordinarily, as O'Donnell (2014) pointed out, corpus analysis would involve the use of concordancing software (e.g. Sketch Engine, Ant Conc). This software allows for the collection of large banks of data (corpora) to be disseminated and analysed for specific speech and grammatical patterns. As Stubbs (2006, p. 17) points out, concordances 'allow us to see patterns and structures at a glance', revealing features that might otherwise have gone unnoticed.

Established English-language corpus resources such as the British National Corpus (2007) have large word counts incorporating a range of text types from different genres and dialects, as well as providing more technical aspects for the grammarian such as formal grammar tags, i.e. POS-tagging. Despite efforts to improve resources in the Welsh language, these are still fairly rudimentary and low in number. A survey of Welsh language corpus resources was undertaken revealing four possibilities as outlined in Table 1. It was determined that a priority for this project was easily accessible and manipulatable data, open source so as to avoid further complexities around data access. It was also important that the text had been POS-tagged to capture the formal grammar choices that realize these functions. This eliminated the majority of resources, as many were simply available in text form only, even if the concordance software itself did have the capacity to handle POS-searches. Given these criteria, it was decided that the CEG corpus would be the most suitable (Ellis et al. 2001).

Corpus Name	Platform/Software	No. words	Features
Welsh Web Corpus	Sketch Engine	50,392,441	Some text metadata available Not POS-tagged
Welsh Web Corpus 2013 (trial version)	Sketch Engine	12,458,397	Some text metadata available Not POS-tagged
Siarad	Downloadable via Bangor University website	460,000	Spoken Welsh only Not POS-tagged Raw data available open-source
CEG - Cronfa Electroneg o Gymraeg (Electronic Database of Welsh	Downloadable via Bangor University website	1,079,032	Extensive text metadata available POS-tagged Raw data available open-source

Table 1: Survey of available Welsh corpus resources as at December 2018

3.3.2 Determining a sample

As shown in Table 1, Ellis etc. state that their CEG corpus consists of almost 1.1million words. While not a large corpus when compared with some of the more established corpora, this still poses a problem of volume for the functional researcher. Firstly, identifying functional features involves interpretation of the text to identify it purpose in the context of the wider text. While advances have been made in the field, this still relies heavily on human interpretation. Furthermore, in the absence of any functional description of Welsh, it is not possible to simply extract instances of formal patterns that we believe might be relevant. For example, looking at noun alone brings back the following results, but not all are the noun groups we need, nor can we be sure that other formal groups aren't also carrying out this function. It is therefore necessary, as O'Donnell (2014) predicted, to select a smaller sample of the larger corpus which would balance out the needs of practicality (manageable sample size) alongside quality (ensuring we can see enough patterns.

To begin the sampling process, the CEG 'metadata' file was downloaded from the project website which gave a breakdown of information regarding each text, including text type, author and year published. From here, it was possible to identify two sub-corpora: fiction and non-fiction texts. Recalling the findings of the lit review in Chapter 2, the sub-corpus of fiction texts was excluded on the basis that they may contain examples of Literary Welsh, a somewhat artificial version of Welsh rarely observed outside religious or literary settings (King 2013).

The sub-corpus of non-fiction texts was then analysed for frequency of text type. This identified 'Press – Scientific', 'Admin – Report', 'Biography', 'Academic' and 'Community' as the five most frequent text-types in the corpus, and therefore it was decided that one text from each would be used as a means to get a variety of texts involved. This was achieved using Excel's randomization function, which selected one text from each text type for analysis.

3.3.3 Building and manually tagging the data sample

The next challenge was devising a way to interrogate the data given the lack of appropriate software to handle both Welsh language texts and the pre-tagged data provided via the CEG researchers. While O'Donnell's (2014) Corpus Tool was built with SFL/SFG analysis in mind, it could not retain the Part-of-Speech tags generated by the CEG corpus. These were deemed important for answering the research questions, therefore it was decided that a manual approach would need to be devised, using Excel to create concordance lines that could be labelled and disseminated for analysis.

The five raw texts along with their tagged counterparts were downloaded from the CEG website under an open source agreement. In doing so, the tagged files had become

corrupted and a Python script was subsequently written to re-format the files for analysis. The output was posted as 'tab separated value' files, which recorded each word, the POS-tag and their lemma (the base word). It also retained the ID for each word, so that any issues could be tracked back to the original text.

3.4 Analysing the data

3.4.1 Identifying grammatical Subjects

Next, it was necessary to attempt to isolate environments in which NGs were most likely to appear. As discussed in Chapter 2, NGs are the most common means of realising a Subject of a clause in English, as Subjects answer the 'What? Or Who of the activity taking place in a given situation. Fontaine also describes them as "*the single most concentrated source of meaning in the clause"*, as it is the position in which all three of Halliday's metafunctions (experiential, interpersonal and textual), conflate (Fontaine 2012). A second practical reason is that Welsh has a grammatical feature that is commonly described as a 'verbal noun'. This is a situation in which an infinitive, e.g. *gyrru* ('to drive') appears as the Complement of a clause with an auxiliary 'Be' verb in initial position, instead of an inflected verb. There has been considerable debate on the nature of this feature in generative circles, with Borsley et al. (2007a) classing these as verb phrases, with Willis (1988) arguing that they are, in fact, noun phrases (i.e. NGs). There has yet to be a functional exploration of this issue, therefore complements were excluded to avoid making a judgement on this highly complex and contentious area of Welsh syntax.

Each text file was then imported into Excel to be read and tagged for Subjects to create a dataset that would, in theory, provide sufficient examples of Welsh NGs to begin a functional analysis of their experiential structure. These were extracted from the tagged file

and transformed into a new Excel file in 'concordance' line fashion, mimicking that of corpus software.

3.4.2 Identifying NGs

Each NG was then analysed for functional features according to Halliday's descriptions of the functions in English. Each concordance line was then centred around the Thing given its core semantic value in the NG. It was then possible to begin quantitatively describing the frequencies of certain functional features, as well as referring to the wider clause or text for further qualitative understanding. The Thing was identified in each case according to

Once identified and tagged, a second python script was written to extract the three types of NGs from the raw tagged files and transforming them into new files along with their identifier and POS-tag. The NGs were centred according to the Thing element. This allowed for concordance-style analysis, with the Thing acting as the key element in context, and the surrounding NG elements present for an investigation into their function and relationship with the Thing. It was decided that, due to the time and scale limitations of the study, investigation would be limited to what Halliday has called 'inter-organism orientation', that is looking at patterns within the group rather than outside of it.

The CL-style concordance layout allows for both quantitative and qualitative research. Frequencies of elements (such as Thing, or formal word classes such as 'nouns') could be quantified through the use of formulae and search functions. This will give an overview of the distribution of patterns within the texts. The analysis will largely be qualitative however, involving the identification of functional features on a case-by-case basis throughout the interrogation of the text, while also being able to identify distinct features and the formal word classes that represent them. This is a corpus-driven approach, meaning that the data

would be analysed after being extracted from the texts, rather than using patterns in the data to answer a hypothesis. This is because the functions of interest were not tagged in the data, rather the data had to be manually analysed to determine if they were present.

Once the expression was centralised around the Thing, the wordings on either side were analysed with respect to experiential functional features identified in Halliday and Matthiessen (2014) and outlined in Chapter 2 Literature Review. The intention was to determine if the following functions were also present in Welsh, and if they appeared in the same position: Deictic, Numerative, Classifier, Epithet and Qualifier. It was important not to impose English NG labels onto Welsh, rather to let the functions come through via the language use. This was also driven by the word choice and placement alone, and not by the POS tag, following Stubbs's warning that meanings do not directly map to grammatical units. The purpose of this exercise was to firstly observe relations between each element, such as its function to specify or to modify the Thing. Secondly, once separated, it allowed for analysis of the paradigmatic choices available within the Welsh NG. In other words, it would demonstrate the various potential ways that a speaker might utilise that specific function, and how their word choices might differ. As well as demonstrating what is possible, it demonstrates what patterns are not. This is the key beneficial feature of this empirical approach over introspection, as the description is driven by the instances in the data alone, allowing us to determine boundaries and identify forms which are not possible.

As with many qualitative methods, this process was iterative. This meant that on closer analysis, errors were identified in the data and corrected (there were four examples of mis-tagged word classes) as well as allowing for a revision of a functional label should further information become available. Furthermore, this study is not interested in incomplete

clauses found in items such as headlines, lists and so on. These often lack the traditional elements of a clause and could give an unrealistic impression of NG structure. Given that the corpus under study includes texts such as press reports and meeting minutes, it is likely that these will be encountered but any NGs found within were excluded.

3.5 Limitations

Given the nature and scale of this thesis, it would not be possible to carry out a comprehensive analysis of all Welsh NGs, nor would it be practical to process large amounts of data. As such, the following limitations have been acknowledged:

Scale

O'Donnell (2014) predicted, it would not be possible to carry out a large-scale analysis of the corpus in the same way that perhaps 'true' CL investigations might prefer. This is because of the considerable amount of manual analysis needed to tag the data with functional labels. However, as Lundquist (2009) points out: small amounts of data can still yield good results, as has been shown by studies such as Nelson (2002) on using the British National Corpus (BNC).

A second limitation is that the small sample size means that it is not possible to test for significance, meaning that the quantitative analysis done in this study will involve descriptive statistics only.

Error

Secondly, the data source (CEG) have already acknowledged that there will be tagging errors within their data, and indeed there were eight instances that were noticed and manually corrected. This is an accepted limitation of CL methodology, and it is believed that the

impact is negligible int his case due to the manual analysis involved. Similarly, manual tagging of a corpus is an inherently introspective and subjective task. While the research was designed such that it could be repeated and verified, it is still entirely possible that manual mis-labelling would occur, and also that another researcher might decide to label some features differently. This error has been reduced to some degree by restricting the scale and scope of the analysis. Furthermore, in making the data available and transparent, and in operationalising the approach taken, these steps can be recreated in future to either verify the findings, or to test against a wider dataset.

3.6 Summary

This chapter has built on the information gleaned in the lit review to put forward a proposal for an empirically-oriented data collection and analysis process. This reflects both the call from Halliday to take on CL, but also calls from researchers pressing for more systematic and unbiased research methods within the field. argued for the importance of empirical data in systemic functional theory and outlined a process by which to apply empirical methods to the creation of a functional description of Welsh NGs.

This chapter has outlined both the rationale for using empirical data alongside CL and SFG analysis, to join theory with method. It has also provided an explanation of the data selection process, sampling etc. It has also demonstrated how the data was analysed in order to answer the specific research questions. How CEG was chosen following a study of current resources available, and the challenges faced with including the data and how this was overcome.

Chapter 4 Functional description of NGs in Welsh: corpus findings, analysis and discussion

4.1 Introduction

This chapter takes the corpus data gathered and analysed as outlined in Chapter 3 Research Methods and extrapolates from it the functional patterns present within Welsh Nominal Groups (NGs). The main output is therefore a proposal for a functional description of the experiential structure of Welsh NGs in Subject position. This will be analysed quantitatively and qualitatively through the SFG framework motivated by, amongst others, Halliday and Matthiessen (2014) and Fontaine (2012).

The chapter will begin by first establishing the environments in which an NG is likely to occur, following the suggestion that Subjects are commonly realized by NGs.

Next, the chapter will analyse these findings and discuss them in terms of what the patterns found, and the frequencies in which they appear, tells us about the nature of referring expressions in Welsh. It will explore each function proposed by Halliday in turn, establishing whether these exist in Welsh and indeed, if so, whether they behave in the same manner. This will be achieved by presenting both the qualitative and quantitative findings of the corpus analysis, resulting in an understanding of the language based on instances as evidence.

Finally, this analysis will close by proposing an order for functional elements in NG in response to both Halliday's proposed ordering of English and Borsley et al's (2007a) ordering of formal elements in Welsh. The intention is to answer whether or not patterns of

specificity can be observed in the NG as it is in English, and propose an order for these elements in Welsh.

4.2 Realization of grammatical Subject

As previously discussed in Chapter 3, before undertaking a manual identification of NG features, it is first necessary to identify conditions in which NGs are most likely to occur in Welsh. A detailed process for this exercise was described in Chapter 3 Research Methods, but in brief it was decided to first identify the Subject of each independent declarative clause, following Fontaine's assertion that Subjects of a clause are most commonly realized by NGs. The intention for this section is therefore to identify if this is the case, and what possible structures have been observed in the dataset.

In total, 766 grammatical Subjects were identified across the sample of five texts. This analysis identified four possible grammatical options for representing the Subject in the dataset, which are outlined in Figure 1 and Table 2.

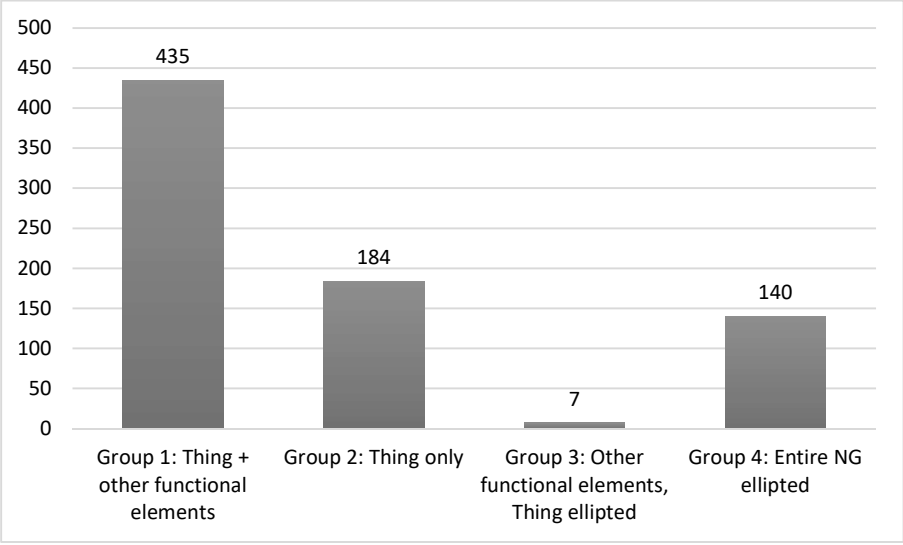

Figure 1: Realization of Grammatical Subject across the dataset (n=766)

Group 1	NG: Thing plus other functional elements, e.g. Deictic, Epithet
Group 2	NG: Thing only
Group 3	NG: Thing ellipted, but other functional elements present
Group 4	Entire NG ellipted

Table 2: Categories of grammatical Subject

As illustrated in Figure 1, in the majority of cases (56.8%), Subjects were represented by NGs comprised of a Thing plus additional modifying functions (**Group 1** Subjects). Examples of these structures can be seen in [1] and [2].

Note: For clarity and consistency, each example is introduced by its English translation. The analysis is then shown in table format below it, with the Welsh NG provided in bold along with the Part-of-Speech (POS) tag that was attributed to each word by the CEG corpus. Finally, the functional elements that were identified manually are captured at the bottom of the table. Additional experiential labels, e.g. Qualifier, have been provided here for clarity, but will be introduced independently later in the chapter.

[1] 'the scientist'

	y	**gwyddonydd**
POS-tag	DefArt	nm
Experiential function	Deictic	Thing

[2] 'a series of portraits of fellow members'

	cyfres	**o bortreadau o aelodau cyd**
POS-tag	nf	prep – npl – prep – npl - nm
Experiential function	Thing	Qualifier

The next most frequent structure was **Group 2** (24.0%), where the Subject was represented solely by a Thing. Examples [3] and [4] show instances of these standalone Things observed in the texts.

[3] 'a strike'

	streic
POS-tag	nf
Experiential function	Thing

[4] 'him'

	ef
POS-tag	pron
Experiential function	Thing

Next, Figure 1 shows that there were 140 cases (18.3%) across the dataset in which the Subject was ellipted (**Group 4**). This means that no NG was observed but that the presence of the Subject was implied by the wider clause. These were coded as 'Ellipted NGs' in the dataset to enable them to be recovered for analysis. Examples **Error! Reference source not found.** and **Error! Reference source not found.** are full clauses taken from the dataset in which the Subject was entirely ellipted.

[5] '<she> has aesthetic and moral experiences'

mae ganddi <hi> brofiadau esthetig a moesol

[6] '<they> succeeded in taking over significant parts of the country that's now known as England'

llwyddasant <hwy> i feddianu rhannau health o'r wlad a adwaenir heddiw fel Lloegr

Qualitative analysis of the co-text of **Group 4** Subjects was carried out in order to understand the nature of these cases and determine their relevance to this study. This analysis showed that all cases were preceded by verbs or verbal groups that were marked for the tense, number and gender of the referred entity in question. For example, in example [6] *llwyddasant* is inflected for the third person plural referent, in this case the Saxons who were introduced in the preceding clause.

Finally, there were a small number of cases (0.9%) in which a Subject was realized by a NG, but that the Thing element had been completely ellipted. These were classed as **Group 3** Subjects for this study. As is shown in example [7], other functional elements were present in the NG to modify and specify the Thing, but the Thing itself was ellipted and recoverable from the other functional elements present.

[7] 'one of <them>'

	un	**ohonynt**	**<hwy>**
POS-tag	card	prep	N/A
Experiential function	Numerative	Selector	<Thing>

Discussion

This initial analysis of the dataset has revealed four different grammatical structures by which Subjects can be realized in Welsh. As might be expected, the most common means of

doing this is through the NG, either by representing the imagined item as a standalone element Thing (i.e **Group 1**) or by accompanying it with other functional elements to further specify and classify the Thing (**Group 2**). What has been revealed, however, is that there exist two other Subject structures in Welsh which have not yet been observed in SFG literature relating to English. That is, that not only is it possible to infer a Subject through elliption of the NG entirely (as in **Group 4**), it is also possible for NGs to represent the Subject with an ellipted Thing (i.e. **Group 3**, example [7]).

The consequences are twofold: firstly, this appears to demonstrate that Welsh can be, at times, a null-Subject language. As Baker (2010) suggested, looking at real data has revealed patterns that we might not have realised otherwise .

Secondly, this initial exercise has allowed for **Group 4** Subjects to be excluded from further analysis as there are no NGs visible to be analysed. Following the qualitative analysis of concordance lines, one possible explanation for these structures is that they are only used when the Subject is immediately recoverable from the text. That is to say, the Subject has been introduced earlier on in another clause and one way of continuing with it is through the use of a verb phrase which is marked for the tense, gender and number of the signalled Subject. Halliday calls such devices 'cohesive devices' as they help build the readability and efficiency of a text. While an interesting observation, and certainly different to the way this is done in some other languages, cohesive features are concerns for the Textual metafunction, and therefore outside the scope of this study. Consequently, all 140 cases of **Group 4** Subjects have been excluded, leaving 626 Subjects (=NGs) for further analysis.

Having now identified three different NG combinations that can realize the grammatical Subject of a clause in Welsh, it is possible to begin a more detailed investigation into the

internal structure and functional patterns observable in NGs. This will commence by first looking at the Thing, given its status as a 'central element' of a NG and the feature that refers to the experiential item in question. This analysis will then be expanded to look at the other functional elements observed within these NGs.

4.3 Thing

As outlined in Chapter 3, the Thing was identified as the core referent in the NG, the function to which all other functions align. This section will detail the quantitative findings of the analysis of Things, namely illustrating the formal means by which a Thing can be represented in Welsh.

4.3.1 Things realized by nouns

The CEG corpus POS-tagger contains a number of sub-categories for nouns. To aid comparison with other word classes, these were provisionally grouped together under the category 'nouns'. Figure 2 demonstrates that, in the overwhelming majority of cases, Things

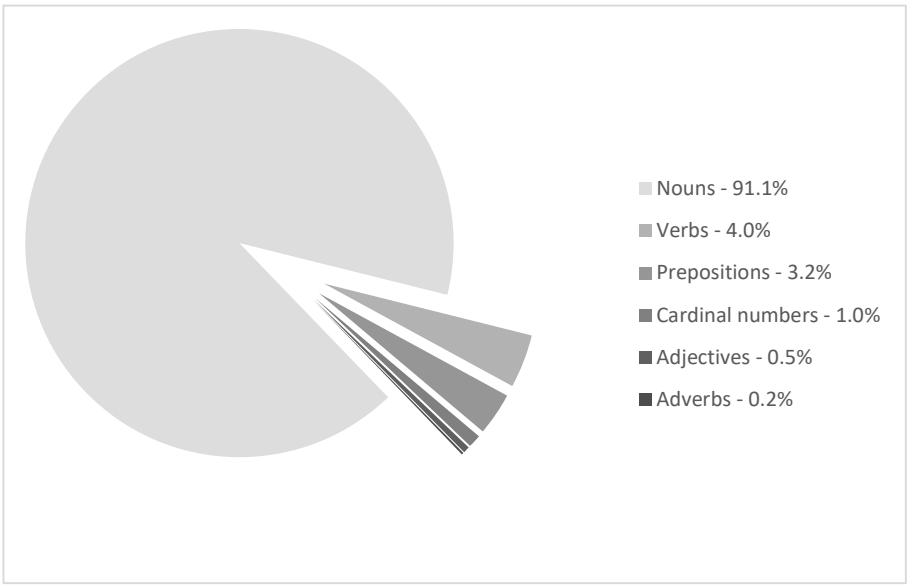

Figure 2: POS-tagged word classes realizing the Thing function in Group 1 and Group 2 Subjects

are realised by nouns in Welsh, representing 91.1% of Things identified in Group 1 & Group 2 Subjects. Use of nouns was higher in Group 1, i.e. Thing plus other elements, than in cases where the Thing was the only element of the NG.

The analogous group 'nouns' was further broken down by POS tag to determine the various subcategories of nouns present (see Figure 3). Again, these were the individual formal classes attributed to the corpus item by CEG's POS-tagging software. This revealed that, in the cases identified, there were six possible noun types realizing the function of Thing. The most populous groups were nouns marked for singular masculine noun (tag = nm), plural noun (tag = npl) and feminine noun (tag = nf). A very small number of cases demonstrated that it was possible for singular nouns to be ambiguous or neutral in gender (tag = nm/f). It was also possible for the Thing to be represented as pronouns (tag = pron) and Demonstrative Pronouns (tag = DemPron), although these were less frequently observed than cases of 'pure' noun.

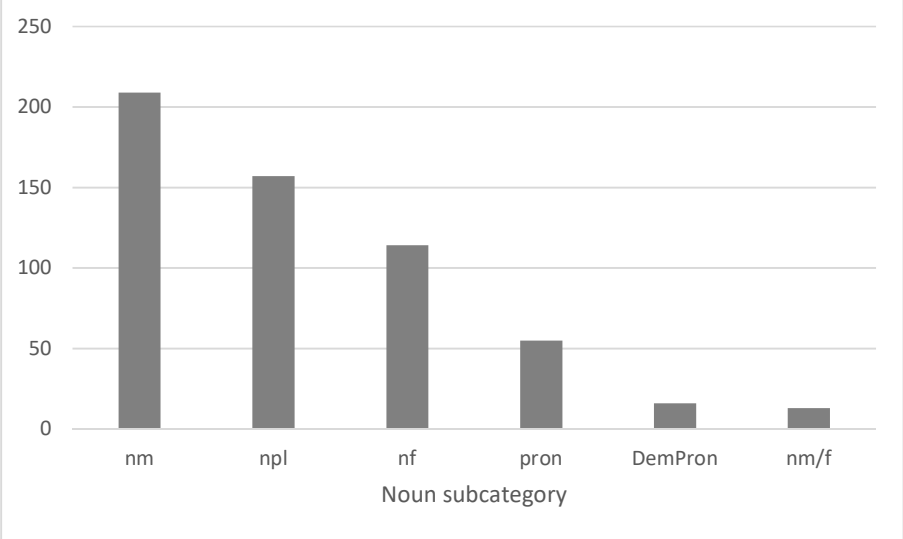

Figure 3: Subcategories of nouns as Things across Group 1 and Group 2 Subjects

4.3.2 Things realized by other word classes

Error! Reference source not found. illustrates that there were instances, representing 8.9% of cases, in which the POS-tagger labelled the Thing as something other than a noun. The first most populous word class after noun for representing the Thing is the verb, representing 4% of cases. Further qualitative analysis of these cases showed that the POS-tagger had been unable to make a distinction between a verb and a nominalised verb, i.e. a gerund or '-ing' form, as in examples [8] and 0.

[8] 'shooting with a gun'

	saethu	**â dryll**
POS-tag	vb	prep – nm
English translation	*shooting*	*with a gun*
Experiential function	Thing	Qualifier

[9] 'significant rearrangement'

	ad-drefnu	**health**
POS-tag	vb	prep – nm
Experiential function	Thing	Qualifier

After verbs, the next most frequent word class in Thing position were prepositions (tag = prep), representing 3.2% of cases in Group 1 & 2. Cardinal numbers (tag=cardn), adjectives (tag = adj) and adverbs (tag = adv) represented the smallest number of cases of realizing the thing, representing 1.0%, 0.5% and 0.2% respectively.

Discussion

Analysis of Thing elements across **Group 1** and **Group 2** has shown that Things in Welsh NGs are most commonly represented by some form of noun. This is consistent with the idea that, in realizing the grammatical Subject of a clause, they are necessarily representing the 'who?' or 'what?' that is participating in the action. The results also confirmed King's (2013) assertion that nouns in Welsh are marked for gender as well as number. This was reflected by the findings displayed in Figure 3, which showed the frequency of nouns in the dataset according to their POS-tag. Next, a key finding was that there were almost 9% of cases in which the Thing was represented by something other than a noun. This is initially surprising as this contradicts Halliday's prediction of the nature of Things and the entities they represent. However, further analysis of individual cases by looking at concordance lines showed that this appears to be an inherent problem of POS-tagging. Examples such as [8] and [9] reveal that nominalisation is present in Welsh, but that the tagger is unable to identify such a feature. This confirms that Welsh has the ability to 'package' processes into nouns so that the processes themselves can participate in another action in a clause. This was not documented in the literature covered in Chapter 2, and proposes an interesting avenue for further exploration outside this study. This feature of nominalisation raises a concern that POS-tagging software is not capable of distinguishing between verbs behaving as the process of a clause, and verbs that have been nominalised. The tagger has identified that the word has a verbal element and has tagged it as a verb accordingly. It demonstrates that looking at instances in context can therefore reveal more than could be analysed automatically.

Finally, in cases where the Thing is indefinite, no other functional items are necessary. This was demonstrated in example [3] where the indefinite nature of *streic* ('a strike') was

signalled by the absence of a definite article (tag = DefArt). This is supported by the descriptions provided by Borsley et al. (2007a), explaining that there is no indefinite article in Welsh.

4.4 Deictic

Thus far, this chapter has focussed on identifying the three NG categories for analysis and providing a description of how Things are represented by formal word classes in the dataset. Next, this section will aim to provide a description for the other experiential functions present in the NG, inspired by descriptions provided by Fontaine (2012) and Halliday and Matthiessen (2014) for English. As **Group 2** consisted only of the Thing element, these cases will be excluded from this analysis. The focus will therefore be on **Group 1** (NG: Thing plus additional functional elements) and **Group 3** (NG: Other functional elements, but Thing ellipted).

Analysis revealed five main categories of function in the NG (other than the Thing): Deictic, Numerative, Epithet, Classifier and Qualifier.

There were 306 NGs that contained a Deictic function in the dataset. These consisted solely of **Group 1** Subjects; there were no cases in which an ellipted Thing was present alongside a Deictic element (i.e. **Group 3**). Furthermore, Deictics were observed in two locations within the NGs: before the Thing, after the Thing, or both. This distribution is shown in Figure 4.

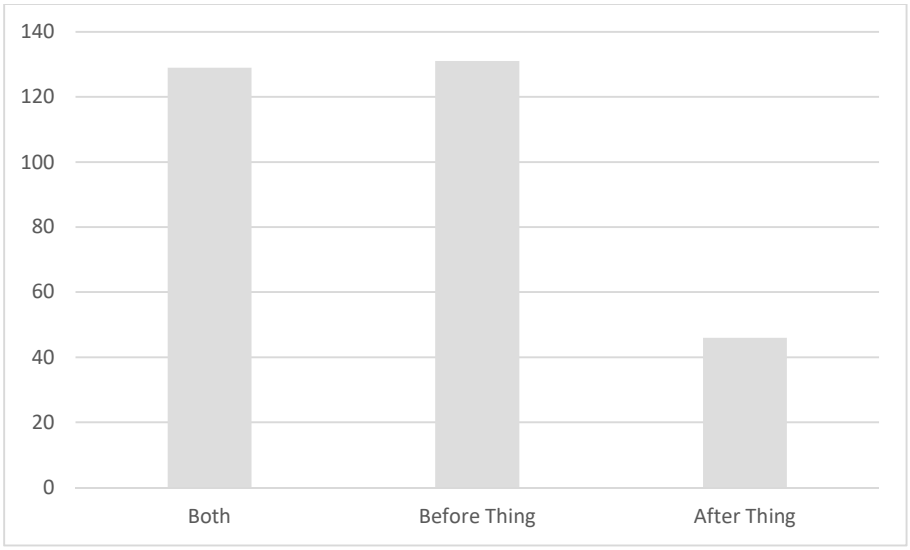

Figure 4: Location of Deictic function in applicable NGs (n=306)

Further analysis was therefore required to determine if the location of the Deictic function had any bearing on its meaning in the NG.

Deictic function before the Thing

As shown in Figure 4, there were 131 NGs in which contained only a Deictic function pre-Thing. These cases were analysed quantitatively to determine the frequency of formal items realizing this function. The vast majority were definite articles (tag = DefArt) representing a form of the word *y*, which translates to the English 'the'. These, in turn, fell into one of three possible representations according to the nature of the words that surrounded them.

Example **Error! Reference source not found.** shows an example where the definite article *y* preceded a word beginning with a consonant. The definite article retains its core form and is unchanged.

[10] 'the job'

	y	**swydd**

POS-tag	DefArt	nm
Experiential function	Primary Deictic	Thing

Example [11], meanwhile, demonstrates how the definite article is transformed to *yr* when preceding a word beginning with a vowel (note that Welsh has two additional vowels: *w* and *y*).

[11] 'the artist'

	yr	**artist**
POS-tag	DefArt	nm/f
Experiential function	Primary Deictic	Thing

Finally, when the definite article follows a word that ends with a vowel as in example [12], that preceding word and the definite article are contracted, resulting in *'r.* The additional 'r' is present, even when the word following it begins with a consonant.

[12] 'five of the parishes'

	pump	**o**	**'r**	**plwyfi**
POS-tag	card	prep	DefArt	npl
Experiential function	Numerative	Selector	Deictic	Thing

After definite articles, the only other word class present in the pre-Thing Deictic category were pronouns (tag=pron). Closer examination of these cases showed that these were all possessive pronouns, a description of each is provided in Table 3. There were no examples of the first-person singular *fy* ('my'), or second-person singular *dy* ('your'). The gender and

number of the pronoun is marked for the gender and number of the possessor, not the Thing (the possession).

Pronoun	Translation	Description
ei	'his'	Third person singular masculine
ei	'her'	Third person singular feminine
eu	'their'	Third person plural masculine/feminine/mixed
ein	'our'	First person plural masculine/feminine/mixed

Table 3: Pronoun types present in the dataset

Deictic functions after the Thing

The next category to be explored are the Deictic functions that were observed after the Thing. Of the 306 NGs affected, 46 contained these features only, i.e. there was no pre-Thing Deictic present. Closer analysis of these cases showed that all were realized by embedded NGs, such as in examples [13], [14] and 0. Note that English translations are approximations provided with literal translations in brackets where applicable.

[13] 'the instrument's form'

	ffurf	**yr**	**offeren**
POS-tag	nf	DefArt	nf
Experiential function	Primary Deictic	Deictic	

[14] 'objects of faith' (lit. 'faith's objects')

	gwrthrychau	**ffydd**
POS-tag	npl	nf
Experiential function	Thing	Deictic

[15] ‘members of the community’ (lit. ‘community’s members’)

	aelodau	**cymuned**
POS-tag	npl	nf
Experiential function	Thing	Deictic

NGs with Deictic elements before and after the Thing

There were 129 NGs in which both the pre-Thing Deictic was present, as well as a Deictic function after the Thing. With one exception, these represented cases in which the first Deictic was realized by a Definite Article, but the post-Thing Deictic were realized by Demonstrative Pronouns or Adverbs. Examples [16] and [17] demonstrate examples of both.

[16] ‘this message’ (lit. ‘the message this’)

	y	**neges**	**hon**
POS-tag	pron	vb	pron
Experiential function	Deictic	Thing	Deictic

[17] ‘that unpleasantness’ (lit. ‘the unpleasantness that’)

	yr	**anhwylder**	**hwnnw**
POS-tag	pron	vb	pron

Experiential function	Deictic	Thing	Deictic

[18] ‘these girls’ (lit. ‘the girls here’)

	yr	**hogiau**	**yma**
POS-tag	pron	vb	pron
Experiential function	Deictic	Thing	Deictic

Finally, as noted, there was one case across this particular subset of the corpus which did not adhere to this pattern. Example [19][19] shows that it is also possible for the first Deictic to be realized by a pronoun and the post-Thing Deictic to be realized by an embedded NG. Note that this also shows an example of a nominalised verb in Thing position.

[19] ‘his self-defence’ (lit. ‘his defending of himself)

	ei	**amddiffyn**	**ei**	**hun**
POS-tag	pron	vb	pron	nf
Experiential function	Deictic	Thing	Deictic	

Discussion

Analysis of this category has revealed features of the Deictic that are quite different from those observed in English (e.g. Halliday, Fontaine). Firstly, the Deictic function was observed in two locations: before the Thing and after it. It was also possible for an NG to contain both. Analysis of features in both locations revealed that they had an implication for meaning, that it was not possible to interchange locations without altering the meaning of the NG.

This suggests that, while both locations carried the role of specifying the Thing as Halliday predicted, this responsibility is split. The Deictic before the Thing can be said to do the 'pointing out'. It specifies, through the definite article or pronoun, that it is a particular entity that is being referred to, and not a general one. The Deictic after the Thing, meanwhile, has the function of locating the item, e.g. here or there. In English, both roles are carried out by this those. Furthermore, where the item is general, this is signalled by the lack of any Deictic function, corroborated by King's assertion that there is no indefinite article in Welsh, as there is in English.

The pre-Thing Deictic bears the responsibility for identifying that the item is specific and not general, i.e. what Halliday calls the 'pointing out function'. A preliminary label of Primary Deictic is therefore proposed for this function, as it first alerts the reader to the specificity of the Thing that will follow. The second position, that after the Thing, has the function of proximity: locating the Thing in context. This can be both the immediate textual context. It is proposed that this is named Secondary Deictic on this basis. It is possible to have both Primary and Secondary Deictic functions in Welsh NGs. In this case the phrase would literally translate to the English 'The cat there' with a similar function to 'That cat'.

4.5 Numerative

Numeratives were observed in two locations in the NGs sampled: before the Primary Deictic and after the Primary Deictic. In all NGs with any Numerative function present (n=58), all appeared before the Thing. Fig [5] demonstrates how these were distributed across the applicable NGs. The most common position (58.6%) was before the Primary Deictic function, followed by post-Deictic position (39.7%). There was one case in which an NG had Numeratives in both locations.

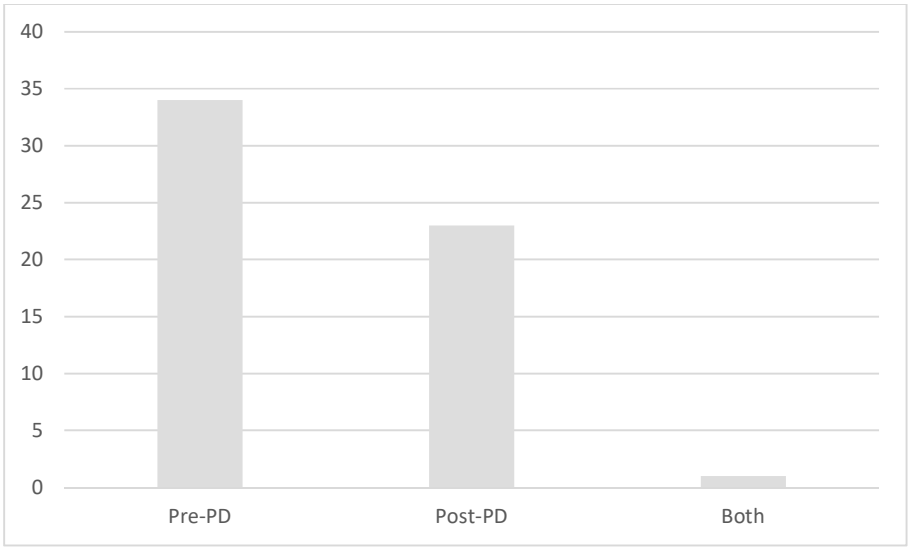

Figure 5: Location of Numeratives in applicable NGs

Numerative before Primary Deictic

Of the 58 NGs that contained Numerative functions, there were 34 instances of those prior to the Primary Deictic. Each instance of Numerative before the Primary Deictic was also followed by an additional functional element: the Selector. This was realised in each case by a form of the preposition 'o' ('of') as shown in example [20]. The Numerative function was the only additional functional element present where the Thing was ellipted (i.e. **Group 4** Subjects).

It is in the Selector element that **Group 1** and **Group 4** Subjects differed. For **Group 1**, all instances of Numeratives before Primary Deictics had a Selector realised by the preposition 'o' ('of') example. For **Group 4**, cases where the Thing was ellipted, this Selector was realised by the preposition 'o' inflected for the gender and number of the ellipted Thing, such as in examples [20] and [21].

[20] '2000 of the men of Cwm Cynon'

	2000	**o**	**ddynion**	**cwm cynon**
POS-tag	card	prep	npl	nplace
Experiential function	Numerative	Selector	Thing	Deictic

[21] 'four of <us>'

	pedwar	**ohonom**	**<ni>**
POS-tag	card	prep	N/A
Experiential function	Numerative	Selector	<Thing>

Numeratives after the Primary Deictic

As shown, there were 23 cases in which only a Numerative after the Primary Deictic was observed. These were realised in formal grammar by a mixture of different word classes as illustrated below:

[22] 'both churches' (lit. 'the two churches')

	'r	**ddwy**	**egwlys**
POS-tag			
Experiential function	Primary Deictic	Numeratative	Thing

[23] 'every happiness' (lit. 'each happiness')

	pob	**llawenydd**
POS-tag		
Experiential function	Numerative	Thing

Both Numeratives before and after the Primary Deictic

There was only once case across this category in which an NG contained both a Numerative and Selector before the Primary Deictic, as well as one directly after. This was example [24] below.

[24] 'one of the four evaluation meetings'

	un	**o**	**'r**	**pedwar**	**cyfarfod**	**gwerthuso**
Experiential function	Numerative	Selector	Primary Deictic	Numerative	Thing	Classifier

Discussion

In Welsh, there appears to be two types of Numerative. The Numerative located at the beginning of the group (i.e. before the Primary Deictic) indicates whether the referent in question represents part of the Thing or not, e.g. 'five trees', 'lots of cheese'. It suggests that the Thing is a whole item, but that the speaker is referring to a part of that whole. This closely resembles what Fontaine (2012b, p. 52) calls the Partitive Determiner, where 'the quantity is being expressed as a selection from a set'. When combined with the Primary Deictic, this becomes more specific, inferring that the 'set' in question is known and not general. It indicates that the set is recoverable in the minds of both the writer and reader, and that the writer is referring to some part of it.

A Numerative after the Primary Deictic serves a slightly different function. In the examples seen in the dataset, it appears that this functions to indicate a quantity of the Thing, as opposed to referring to part of it. This distinction can be seen in example [24], where the

combination of both types of Numerative demonstrates their differing functions. In this case, there are four of the Thing (that is, four meetings) but the speaker refers to one in particular. Here both the partitive function of the first Numerative, and the quantitative function of the second Numerative, combine to make it clear which entity is being referred to. In English, this distinction is less clear cut, as Fontaine points out that the Selector option ('of') is optional, but as can be seen in Welsh, both Numerative types serve very specific functions. It is therefore open to discussion whether this is the case in English, and whether both functions conflate on the same location.

4.6 Epithets

As with both the Deictic and Numerative functions previously covered, it was shown that the Epithet function can also be observed in two places in the NG: before the Thing, and after it. Figure 6: locations of Epithets in applicable NGs illustrates the frequencies of this occurrence across the 74 NGs that contained an Epithet function.

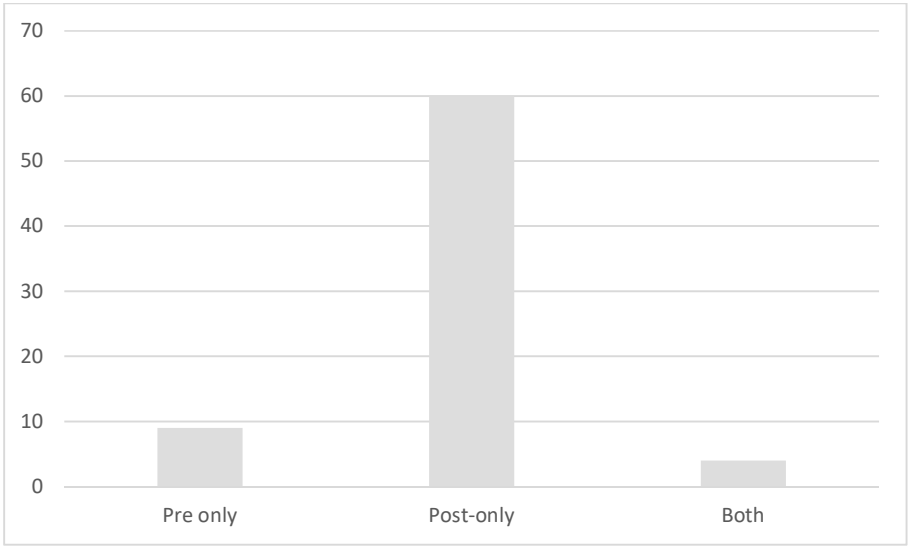

Figure 6: locations of Epithets in applicable NGs

This demonstrates that, in the majority of cases (82.2%), the Epithet only appears after the Thing.

Epithets before the Thing

There were only a small number of cases in which there was an Epithet placed before the Thing. These were located directly before the Thing element, i.e. there were no other functional elements between them. A frequency analysis of all pre-Thing Epithets in the dataset (this included instances where both Epithet locations were populated) was carried out to understand if there were predictable patterns to the pre-Thing Epithets. Table 5 demonstrates these frequencies.

Word	POS-tag	Translation	Frequency in text
hen	adj	'old'	5
prif (inc. brif)	adj	'main', 'primary' or 'chief'	5
union	adj	'exact'	1
math	nm/f	'such'	1
cryn	adj	'barely'	1
bron	adj	'almost' or 'nearly'	1

Table 4: word frequencies within the pre-Thing Epithet category

As can be seen from this analysis, the most common words realising this function were *prif* ('main') or *hen* ('old'). This suggests that there is not much variety and choice in what can precede the Thing in terms of Epithets.

Epithets after the Thing

Analysis of this group demonstrated, firstly, that this was a much more common occurrence in the data set, compared with pre-Thing (again, this includes instances of both). Of these

the majority were realized by adjectives or adjectival groups, with the latter representing cases where the adjective had been intensified as in example

[25] 'a very winding journey'

	taith	**droellog**	**iawn**
POS-tag	nf	adj	adv
Experiential function	Thing	Epithet	

4.7 Classifiers

The next function addressed was the Classifier. This function appeared in just one location in the NG: directly after the Thing, but before the Epithet (if applicable). In total, there were 85 NGs that contained a Classifier. While a qualitative examination of these cases showed that the most frequent means of realising a classifier was via a noun (i.e. tag = nm, nf, nm/f or npl) or an adjective as in example [26]. There was one case in which a Classifier was realized by an embedded NG.

[26] 'South Wales's communications officer'

	swyddog	**cyswllt**	**de-orllewin Cymru**
POS-tag	nm	nm	adj – npl
Experiential function	Thing	Classifier	Deictic

4.8 Qualifiers

The final category of functional element detected in the dataset is Qualifiers. In all 62 cases in which a Qualifier was identified (8.1% of NGs in the dataset). All appeared at the final location in the NG. This was also the case where the Thing was followed by an Epithet, Classifier or Secondary Deictic (or a combination of all three), as in example [27] (only functional elements shown).

[27] ‘an agressive brigade of coal miners’

	brigâd	**ymladdgar**	**o lowyr**
Experiential function	Thing	Epithet	Quallifier

Qualifiers were further categorised by their composition, revealing that all were instances of embedding. For this reason, POS-tagging can be disregarded at this stage. Instead, they appear to be phrases or units that were embedded at this location. Examples [28] and [29] demonstrate this.

[28] ‘the name given to a community’

	yr	**new**	**a roddid ar gymuned**
Experiential function	Primary Deictic	Thing	Qualifier

[29] ‘the church in Llanddewi Brefi’

	yr	**eglwys**	**yn Llanddewi Brefi**
Experiential function	Primary Deictic	Thing	Qualifier

Discussion

When referring to English, Halliday and Matthiessen stated that the Qualifier element follows the Thing and that it bears the function of characterizing the Thing. In Welsh, only the latter definition is true, as the Thing element can be followed by Epithets and Classifiers,

unlike its English counterpart. The Qualifier in Welsh can therefore be seen as providing further contextual information by which to define the entire NG thus far.

4.9 Ordering of functional elements in the NG

As a result of the analysis completed above, it is possible to describe the potential of the Welsh nominal group thus:

Numerative	Selector	Primary Deictic	Numerative	Epithet	Thing	Epithet	Classifier	Secondary Deictic	Qualifier

This is evidently a very different structure to that observed in English, and subsequently requires further analysis of what this pattern tells us about the experiential nature of referring items in Welsh (Halliday and Matthiessen 2014, pp. 380 - 381). Here, it is not possible to see the same 'left to right' cline of decreasing specificity that Halliday and Matthiessen documented in English. Instead, it could be said that the NG becomes more specific as it radiates outwards from the Thing. In turn, the closer to the centre of an NG, the more 'permanent' the attribute.

4.10 Conclusion

This chapter has been concerned with the primary objective of this study: that is to provide a functional description of the experiential nature of Nominal Groups in Welsh. This exercise began by first identifying the conditions which would yield suitable NGs for analysis, establishing that the Subject is represented by NGs in the vast majority of cases, but that it is also possible to ellipt the NG entirely where the referent is known through context or has already been introduced. From here, it was possible to analyse quantitatively and qualitatively 766 NGs in Subject position: some realized as Thing only (Group 2), Thing plus

other functions (Group 1) and rare instances in which the Thing is ellipted but an NG is still present. Six functions were identified in total, and described in detail: the Thing, Deictic, Numerative (and Selector), Epithet, Classifier and Qualifier. These were the same functions that were observed in Halliday's description of English, suggesting that in the case of Welsh, at least, both languages share the same functions. Where they differ, however, is in their location within the NG.

As Fontaine (2007) predicted, there is no reason to expect NG structures to be identical across languages. This has certainly been evidenced through this functional description. Not only have functional elements been shown to appear in different locations to that of the NG, for example the Classifier, they have also been shown to be split across locations, such as is the case for Deictics. It was possible to establish that Welsh utilises two sub-categories of Deictic, leading to the requirement for additional labels. These were the Primary and Secondary Deictic, respectively, where it has been suggested that the Primary Deictic is responsible for the 'pointing out' function that Halliday characterises of the Deictic. The Secondary Deictic bears greater responsibility for locating the item in context or space. Furthermore, it was established that where the possessor was known to the speaker and audience, this was signalled in the Primary Deictic position by a demonstrative pronoun, e.g. ein ('our'). Where the possessor has not yet been introduced in the text, or are outside the immediate context, they are signalled in the Secondary Deictic through use of an embedded NG. It was also demonstrated that the placements for Classifiers, Epithets and Numeratives are different in Welsh, and that some functions can be placed before, after or around other features, but that these differences in locations have subtle implications for meaning.

Finally, a template was proposed for NGs in Welsh, demonstrating the meaning potential of this functional unit. It has been suggested that in Welsh, the Thing is at the core of the NG, and functions of further classification, modification and specification radiate from the centre of the Thing, with the functions with the most capacity for doing so being furthest away from the core of the NG. For example, the Classifier was found closest to the Thing, immediately classifying the type of thing in mind. Meanwhile the Qualifier, which provides additional contextual information to help identify the Thing sits the furthest from the NG's 'core'. This can be contrasted with the NG in English according to Halliday, where it is said that things progress along a cline of less to more specific.

Chapter 5 Conclusion

5.1 Introduction

This thesis has proposed a functional description of Welsh referring expressions realized via the Nominal Group. It has done this through deploying empirical data collection and dissemination methods, as suggested by Hunston (2013) and Stubbs (2006). It has also sought to explain how the different functions operate to create a coherent referring expression, as well as detailing the formal word classes that traditionally realize them. The intention was to propose a pattern within Welsh NGs that reflected function rather than form, to enhance patterns previously described via Transformational-Grammar resulting in a more 'triangulated' approach to our understanding of language (Angouri 2010, p. 34).

It has shown that Subjects in Welsh can take on four forms: NGs which contain the Thing alone; NGs with additional functional items to specify the Thing; NGs in which the Thing itself was ellipted but additional functional elements remained; and finally demonstrated that it is possible to ellipt the NG altogether, with the Subject recoverable from context or a finite verb that precedes it.

Things can be realized by a number of formal word classes, but are most commonly done so via types of noun. Where alternate cases existed, such as verbs or adjectives, these reflected some of the flaws in annotation based on formal word classes, as automatic taggers are often unable to differentiate between true verbs and those that have been nominalised. Welsh Things have the additional complexity of signalling gender as well as number, and these characteristics have consequences for the functional elements that surrounded them.

Like in English, six functional elements were identified: Thing, Deictic, Numerative, Classifer, Epithet and Qualifier. Unlike English, however, these were less clear cut, and, with the

exception of the Qualifier and Classifier, could have forms that appeared before and/or after the Thing as required. Welsh signals specificity through the use of the Primary Deictic but relies upon a Deictic function after the Thing to denote contextual proximity, whereas in English this conflates in one location, such as in 'this' or 'those'. These function to specify the Thing in the consciousness of the audience, as well as indicating if the thing is recoverable from context or co-text. The Deictic also has implications for possession. Pronouns can realize the Primary Deictic function, indicating that the Thing in question belongs to a known person. If the possessor is not known, or is being introduced during the process of referring/NG, then it appears in the Secondary Deictic position.

The Numerative function appears before the Thing in two positions: the first to indicate part of a whole, along with a selector function 'o' of. The second is more definite, defining that there are multiples of the Thing as opposed to representing part of a list of Things. The Numerative function has shown to be the only examples in which the Thing can be ellipted, although the preposition in the Selector function is modified and marked for the gender and number of the ellipted Thing, rendering it recoverable to the reader.

Next, the modifying functions of Epithet and Classifier. These were distinguishable by their ability to be intensified, and indeed there were a number of instances in which intensification occurred. Epithets most commonly occurred after the Thing and served to furnish it with descriptive detail. These were commonly realized in form adjectives. There were specific instances in which the Epithet directly preceded the Thing, but these were limited to a short selection of adjectives. Classifiers were less frequently observed because the nature of Welsh meant that this function was absorbed by the Secondary Deictic. Examples of Classifier, Epithet and Secondary Deictic together were rare, but the examples

identified corroborates that these were separate functions, albeit with overlapping features. this means that Classification and Possession (via Deictic) is sometimes ambiguous in Welsh.

The final function identified was the Qualifier. Unlike English, the Qualifier does not directly follow the Thing, but does sit at the final position in the NG, providing contextual information usually through embedded groups etc to further specify the Thing and the elements that are packaged closely to it.

The consequences for these findings reveals that the ordering of functions in Welsh is different to that proposed by Halliday. This is to be expected given the differences in languages, but serves as a reminder that, as Caffarel proposed, further languages and data can help guide the theory, moving it away from an Anglo-centric theory simply adapted for other languages. The ordering suggests that the most general item – the Thing – sits at the core of the NG. Next, there is an optional 'sandwich' of modification around it, providing further classifying information such as the type of Thing or descriptive characteristics of it. This is more specific than the Thing would be alone. Next come the Deictic functions – primary on one side and Secondary on the other – which, if present, serve to indicate that the Thing is specific, either a shared referent according to co-text and context (i.e. it is not another thing, it is one specific instance of the thing in question which I want you to be able to recover). The Qualifier provides further specificity and the Numerative is more specific again, indicating that the thing is part of a whole. On this basis, Welsh challenges Halliday's version of ordering in English, with the less specific at the centre, and optional, specifying elements moving from the centre out, with the most specific at the outer edges.

5.2 Potential areas for future research

It is hoped that this study has demonstrated the value of including minority languages, and indeed any language, within the SFG canon. This study focussed on one element – the NG – but as indicated early on, these are some of the finer details – the 'little things' – upon which bigger analyses can be carried out.

Firstly, there is the potential for further exploration of the Welsh language itself. As a language with Verb-Subject-Object order, it would be enlightening to explore the nature of other groups and functions as a way of expanding our understanding of how meaning potential is created. One particular domain would be the question of the 'verb-noun' which was briefly introduced in Chapter2 Literature Review. As discussed, this is a function of Welsh grammar in which the property of the element is ambiguous. While attempts have been made to caegorise these (e.g. Willis 1988) these have, again, been limited to the generative framework. A functional analysis of these features may reveal more about the ways in which they function, and also challenge our existing understanding of how processes are conveyed in language.

Secondly, this study focussed on what Halliday has termed the 'inter-orientation' of NGs, that is the functions and patterns within the boundaries of the NG itself. Looking beyond the NG, that is too look at the role it plays in the wider discourse, could take the study of Welsh NG away from pure theory and into the realms of the sociolinguistic. This naturally has implications for the other metafunctions that were not explored here, namely how the language constitutes the text (Textual metafunction) and how language choices reflect the relations between the writer and their audience (Interpersonal metafunction). The former approach could draw upon the wide variety of texts available within CEG to explore how

patterns within NGs construe the text register, e.g. fiction versus non-fiction, academic versus newspaper reports. It could be beneficial to explore if this is consistent with findings in Welsh. There is also potential for an applied linguistics approach, particularly in the realms of Welsh-language teaching. This is a growing industry, and following successes of studies into English and teaching, there could be avenues for exploration here that would help us understand how Welsh is learned.

Finally, as indicated in Chapter 3 Research Methods, the nature of this investigation has necessarily been limited by both time and scale. As O'Donnell (2014) predicted, applying SFG to corpus-style analysis is necessarily time-consuming, mainly due to the amount of human effort involved in inputting SFG tags. A larger scale operation might therefore garner more wide-ranging results.

5.3 Contribution to knowledge

To conclude, this study has proposed a functional description in a language that has yet to have been included to any great extent within the SFG framework. In doing so, it has offered further insights into some of the fundamental concepts surrounding NGs and their meaning potential, following in the footsteps of scholars such as Caffarel (2006) and Moalla (2018) in expanding our knowledge of language out from the current hubs of study English and Chinese, where hubs of SFG and SFL activity has already been carried out. The analysis carried out has also both supplemented and challenged existing research into Welsh syntax via the generative tradition, demonstrating the power of SFG to incorporate context and function to discussions that otherwise look no further than the formal choices that represent them. Finally, it has responded to calls from scholars to move away from introspection as the primary means of developing descriptions of grammar. In adopting a

joint-approach, that is SFG theory with corpus linguistics methodology, it has reinforced suggestions proposed by Hunston, Thomas and O'Donnell, and demonstrates that the potential benefits outweigh many of the perceived complexities. This has potential exciting consequences for our understanding of minority languages.

References

Angouri, J. 2010. Quantitative, qualitative or both? Combining methods in linguistic research. In: Litosseliti, L. ed. *Research Methods in Linguistics*. London: Continuum, pp. 29–45.

Baker, P. 2010. Corpus methods in linguistics. In: Litosseliti, L. ed. *Research Methods in Linguistics*. London: Continuum, pp. 93–113.

Berry, M. 1975. *Introduction to Systemic Linguistics: 1 Structures and Systems*. London: Batsford.

Bloor, T. and Bloor, M. 2004. *The Functional Analysis of English: A Hallidayan Approach*. 2nd ed. London: Arnold.

Borsley, R.D. et al. 2007a. Noun phrases. In: *The Syntax of Welsh*. Cambridge: Cambridge University Press (Cambridge Syntax Guides), pp. 152–197. doi: 10.1017/CBO9780511486227.006.

Borsley, R.D. et al. 2007b. Welsh as a VSO language. In: *The Syntax of Welsh*. Cambridge: Cambridge University Press (Cambridge Syntax Guides), pp. 338–366. Available at: 10.1017/CBO9780511486227.011.

Caffarel, A. 2006. *A Systemic Functional Grammar of French: From Grammar to Discourse*. London: Continuum.

Chan, W.H. and Nicoladis, E. 2010. Predicting two Mandarin-English bilingual children's first 50 words: Effects of frequency and relative exposure in the input. *International Journal of Bilingualism* 14(2), pp. 237–270.

Chomsky, N. 1965. *Aspects on the theory of grammar*. Cambridge, Massachusetts: MIT Press.

Ellis, N.C. et al. 2001. Cronfa Electroneg o Gymraeg (CEG): A 1 million word lexical database and frequency count for Welsh. Available at: https://www.bangor.ac.uk/canolfanbedwyr/ceg.php.en [Accessed: 1 August 2018].

Fawcett, R. 2000. *A Theory of Syntax for Systemic Functional Linguistics*. Amsterdam/Philadelphia: John Benjamins.

Fontaine, L. 2007. The variability of referring expressions. In: *Lacus Forum XXXIII*. Toronto: University of Toronto, pp. 159–170. Available at: http://www.lacus.org/volumes/33.

Fontaine, L. 2012a. Introduction to functional grammatical analysis. In: *Analysing English Grammar: A Systemic Functional Introduction*. Cambridge: Cambridge University Press, pp. 1–15.

Fontaine, L. 2012b. The grammar of things: the nominal group. In: *Analysing English Grammar: A Systemic Functional Introduction*. Cambridge: Cambridge University Press, pp. 44–70.

Fontaine, L. [no date]. The Noun, Grammar and Context. In: Peng, A. X. and Webster, J. eds. *The Making of Meaning: Grammar, Society and Consciousness*. London: Bloomsbury

Gibson, E. and Fedorenko, E. 2013. The need for quantitative methods in syntax and semantics research semantics research. *Language & Cognitive Processes* 0965(1–2), pp. 88–124. doi: 10.1080/01690965.2010.515080.

Halliday, M.A.. and Matthiessen, C. 2014. *Halliday's Introduction to Functional Grammar*. 4th ed. Abingdon: Routledge.

Hannahs, S.J. and Tallerman, M. 2006. At the interface: Selection of the Welsh definite article. *Linguistics* 44(4), pp. 781–816. doi: 10.1515/LING.2006.025.

Hunston, S. 2013. Systemic functional linguistics, corpus linguistics, and the ideology of science. *Text and Talk* 33(4–5), pp. 617–640. doi: 10.1515/text-2013-0028.

Jones, D.E. et al. 2016. Improving Enterprise Wide Search in Large Engineering Multinationals: A Linguistic Comparison of the Structures of Internet-Search and Enterprise-Search Queries.

King, G. 2003. *Modern Welsh: A Comprehensive Grammar*. 2nd ed. London: Routledge.

Levon, E. 2010. Organizing and processing your data: the nuts and bolts of quantitative analysis. In: Litosseliti, L. ed. *Research Methods in Linguistics*. London: Continuum, pp. 68–92.

Lundquist, H. 2009. *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.

Moalla, D. 2018. Nominal Groups in Arabic and English: Experiential differences and effect on Tunisian EFL learners' translations. In: Sellami-Baklouti, A. and Fontaine, L. eds. *Perspectives from Systemic Functional Linguistics*. Abingdon: Routledge, pp. 288–314.

Mwinlaaru, I.N. and Xuan, W.W. 2016. A survey of studies in systemic functional language description and typology. *Functional Linguistics* 3(8). doi: 10.1186/s40554-016-0030-4.

Neale, A. 2006. Matching corpus data and system networks: using corpora to modify and extend the system networks for Transitivity in English. In: Thompson, G. and Hunston, S. eds. *System and corpus: exploring connections*. London: Equinox, pp. 143–163.

O'Donnell, M. 2014. Systemic Funcional Linguistics and Corpus Linguistics: Interconnections and current state. In: Yan, F. and Webster, J. J. eds. *Developing Systemic Functional Linguistics: Theory and Application*. Sheffield: Equinox, pp. 345–369.

Sampson, G. 2001. *Empirical Linguistics*. London: Continuum.

Stubbs, M. 2006. Corpus analysis: the state of the art and three types of unanswered questions. In: Thompson, G. and Hunston, S. eds. *System and corpus: exploring connections*. London: Equinox, pp. 15–36.

The British National Corpus. 2007. Available at: http://www.natcorp.ox.ac.uk/ [Accessed: 15 January 2019].

Thompson, G. 2014. *Introducing Functional Grammar*. Abingdon: Routledge.

Welsh Government 2018. Annual Population Survey estimates of persons aged 3 and over who say they can speak Welsh by local authority and measure. Available at: https://statswales.gov.wales/Catalogue/Welsh-Language/annualpopulationsurveyestimatesofpersonsaged3andoverwhosaytheycanspeakwelsh-by-localauthority-measure.

Willis, P. 1988. Is the Welsh verbal noun a verb or a noun? *WORD* 39(3), pp. 201–224. doi: 10.1080/00437956.1988.11435790.

Appendices

The data for this study is supplied via an attached USB drive.