Parsimonious Vole
A Systemic Functional Parser for English

Universität Bremen

Eugeniu Costetchi

Supervisor: Prof. John Bateman

Advisor: Dr. Eric Ras

Faculty 10: Linguistics and Literary Studies
University of Bremen

This dissertation is submitted for the degree of
Doctor of Philosophy

June 2018

I would like to dedicate this thesis to my loving parents . . .

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

<div align="right">

Eugeniu Costetchi

June 2018

</div>

Acknowledgements

And I would like to acknowledge ...

Abstract

This is where you write your abstract ...

Table of contents

List of figures

List of tables

List of definitions

Chapter 1

The systemic functional theory of grammar

Any description of language requires a theory that provides the frame, scope and the necessary concepts. Having a solid theory of grammar contributes to explaining what language is and how it works. It also frames how language is ought yo be analysed by either human or machines.

In his seminal paper Halliday (1961b) addresses the ardent need of the time for a general theory of language and partially answers the proposal for a universal theory of language. He sets out what was known at the time as Scale and Category Grammar. In such a model *units* are set up to account for for pieces of language which carry grammatical patterns. They are seen as arranged on a hierarchical *rank* scale of words, groups and clauses. These and other foundational concepts are covered in the first part of this Chapter.

There are two variants of Systemic Functional Grammars: the *Sydney Grammar* started in 1961 by Halliday (2002) and the *Cardiff Grammar* proposed by Fawcett (2008) which is a simplification and an extension of the Sydney Grammar. To understand the underlying common motives and how they are different we shall start looking at their theories of grammar. They also have quite different historical developments.

Sydney and Cardiff grammars have been formalised to the point where they could be computationally applied to natural language generation. They have been implemented in PENMAN (Mann 1983b; Penman Project 1989) and respectively COMMUNAL projects (Fawcett 1990). Both versions of SF grammars have been used predominantly for English and implementations of for other languages are also available. The major component of PENMAN is a computer model of Halliday's SF grammar described by Mann & Matthiessen (February 1983), Matthiessen & Bateman (1991), (Matthiessen

1995) and others. COMMUNAL is the computer implementation of Cardiff grammar described by Fawcett (1988a), Fawcett (1993) and others.

This chapter first sets out the basic organisational dimensions for each of the theories and then discusses comparatively Halliday's (Halliday 2002) and Fawcett's (Fawcett 2000) versions of SFL.

1.1   A word on wording

Before going into deeper discussion I first make terminological clarifications on the terms: grammar, grammatics, syntax, semantics and lexicogrammar. I start with a definitions adopted in "mainstream" generative linguistics and then present how the same terms are discussed in systemic functional linguistics.

Radford, a generative linguist, in the "Minimalist Introduction to Syntax" (1997), starts with a description of grammar as a field of study, which, in his words, is traditionally subdivided into two inter-related areas of study: syntax and morphology.

Definition 1.1.1 (Morphology (Radford))**.**   Morphology is the study of how words are formed out of smaller units (traditionally called morphemes) (Radford 1997: 1).

Definition 1.1.2 (Syntax (Radford))**.**   Syntax is the study of how words can be combined together to form phrases and sentences. (Radford 1997: 1)

Halliday, in the context of *rank* scale discussion (see Definition 1.2.1 and 1.2.2), refers to the traditional meaning of syntax as the *grammar above the word* and to morphology as *grammar below the word* (Halliday 2002: 51). Such a distinction, he states, has no theoretical status and is deemed as unnecessary distinction. Halliday adopts this position to motivate the architecture of grammar he was developing and is inherited from his precursor, Firth, as he puts it:

> . . . the distinction between morphology and syntax is no longer useful or convenient in descriptive linguistics. (Firth 1957: 14)

Radford adds that, traditionally, grammar is not only concerned with the principles governing formation of words, phrases and sentences but also with principles governing their interpretation. Therefore *structural aspects of meaning* are said to be also a part of grammar.

Definition 1.1.3 (Grammar (Radford))**.**   [Grammar is] the study of the principles which govern the formation and interpretation of words, phrases and sentences. (Radford 1997: 1)

Interestingly enough, the Definition 1.1.3 makes no mention at all to the lexicon. This is because the formal grammars focus primarily on unit classes and how they are accommodated in various structures and so in formal linguistics the lexicon is often disconnected from the grammar. The systemic grammar, on the other hand, along with formal descriptions of grammatical categories and structures, includes the lexicon as part of grammar to form a *lexicogrammar*. At this point I have to mention that systemic functional grammar is not the only lexicalised one and there are others taking the same approach such as Lexical Functional Grammar (LFG), Head Phrase Structure Grammar (HPSG), Combinatory Categorial Grammar (CCG) and others.

Another important aspect to notice is that the grammar is defined as a field of study rather than a set of rules. De divergence in perspective on the subject led Halliday, since his early papers, to become conscious the difference between a study of a phenomenon with the phenomenon itself. By analogy to language as phenomenon and linguistics as the study of the phenomenon, discussed in (Halliday 1997), Halliday adopts the same wording for *grammar* as phenomenon and *grammatics* as the study of grammar; the same distinction holds for *syntax* and *syntactics*.

Definition 1.1.4 (Grammatics (Halliday))**.** Grammatics is a theory for explaining grammar (Halliday 2002: 369)

Moravcsik, another generative linguist, stresses the same distinction, in her "An introduction to syntax" (Moravcsik 2006), and presents two ways in which the word *syntax* is used in the literature: (a) in reference to a particular aspect of grammatical structure and (b) in reference to a sub-field of descriptive linguistics that describes this aspect of grammar. In her words:

> . . . syntax describes the selection and order of words that make well-formed sentences and it does so in as general a manner as possible so as to bring out similarities among different sentences of the same language and different languages and render them explainable. . . . syntax rules also need to account for the relationship between string of word meanings and the entire sentence meaning, on one hand, and relationship between strings of word forms and the entire sentential phonetic form, on the other hand. (Moravcsik 2006: 25)

In her definition of grammar she includes the lexicon and semantics which is a somewhat more explicit statement than Radford's *interpretation.* She is also getting, in Definition 1.1.5, somewhat closer to what grammar stands for in SFL - Definition 1.1.6.

Definition 1.1.5 (Grammar (Moravcsik))**.** ... maximally general analytic descriptions, provided by descriptive linguistics, [are] called grammars. A grammar has five components: phonology (or, depending on the medium, its correspondent e.g. morphology), lexicon, syntax and semantics(Moravcsik 2006: 24–25).

Definition 1.1.6 (Grammar (Halliday))**.** To Halliday, lexico-grammar, or for short,simply grammar is a part of language and it means the wording system - the "lexical-grammatical stratum of natural language as traditionally understood, comprising its syntax, vocabulary together with any morphology the language may display [...]" (Halliday 2002: 369).

The last point I want to mention is the approach to semantics. Formal grammars aim to account for the realisation variations, that is formation of words, phrases and sentences along with their arrangements and mention of semantics is often restricted to what may be termed the *formal aspect of meaning.*

By contrast, a systemic grammar is a functional grammar, which means (among other things) that it is semantically motivated, i.e. "natural". So the fundamental distinctions between formal and functional grammars is the semantic basis for explanations of structure.

Also, in SFL, the meaning is being approached from a semiotic perspective, placing the linguistic semantics in perspective with the linguistic expression and the real world situation. In this respect, Lemke (1993) offers a well formulated theoretical foundation that "human communities are eco-social systems that persist in time through ongoing exchange with their environment; and the same holds true for any of their sub-subsystems [...]" including language. The social practices constituting such systems are both material and semiotic, with a constant dynamic interplay between the two. (Halliday 2002: 387)

To Halliday, the term *semiotic* accounts for an orientation towards meaning rather than sign. In other words, the interaction is between *the practice of doing and the practice of meaning.* As the two sets of practices are strongly coupled, Lemke points out that there is a high degree of redundancy between the *material-semiotic interplay.* And it perfectly resonates with Firth's idea of *mutual expectancy* between the text and the situation. This idea of interplay is incorporated in SFL as *language stratification* and is graphically represented in Figure 1.1.

Having that said, the stratification axis is a useful dimension to relate the formal and the systemic functional grammars. This is also an instrument employed by Hjelmslev (Taverniers 2011).
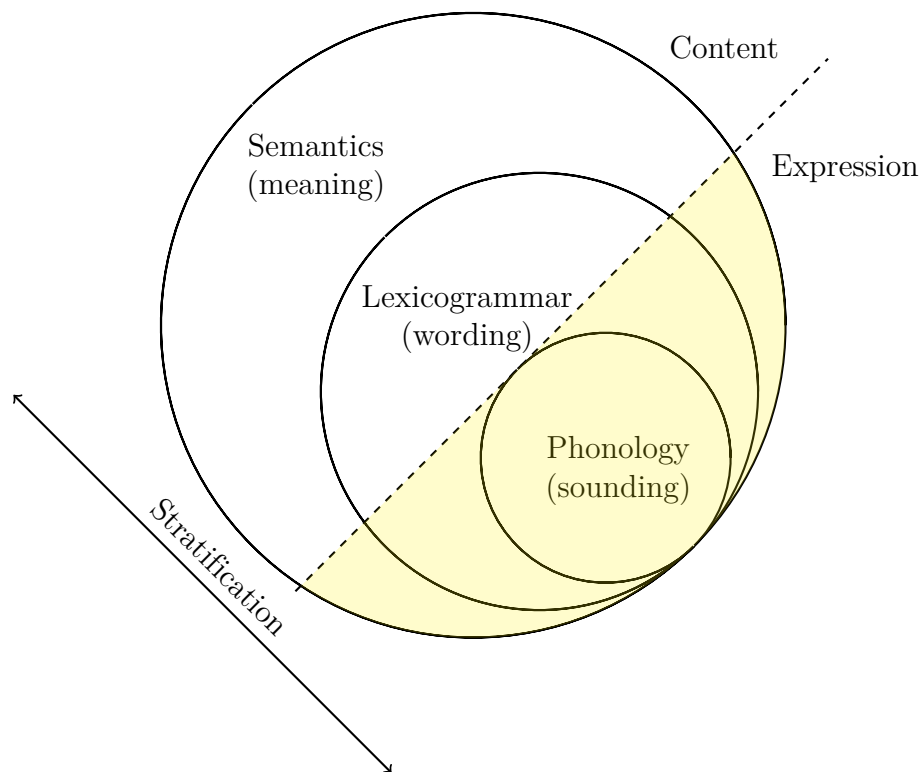
Fig. 1.1 The levels of abstraction along the realisation axis

The SFL model defines language as a resource organised into three strata: phonology (sounding), lexicogrammar (wording) and semantics (meaning). Each is defined according to its level of abstraction on the realisation axis. The realisation axis is divided into two planes: the expression and the content planes. Although debate about the precise division continues, for current purpose it is sufficient to see the first stratum (i.e. phonology/morphology) belongs to the *expression plane* and the last two (lexicogrammar and semantics) belong to the *content plane*. In this context, the formal grammar could be localised entirely within the expression plane, including the phonology/morphology, syntax, lexicon while formal semantics, stripped of any explanations in terms of the meaning potential, belongs in the content plane.

1.2   Sydney theory of grammar

I start introducing the terms of SFL theory with the Sydney grammar as this is in accordance with the historical development originating with Halliday (2002) defining the categories of the theory fo grammar. He proposes four fundamental categories:

unit, *structure*, *class* and *system*. Each of these categories is logically derivable from and related to the other ones in a way that they mutually define each other. These categories relate to each other on three scales of abstraction: *rank, exponence, delicacy.* Halliday also uses three scale types: *hierarchy, taxonomy* and *cline.*

Definition 1.2.1 (Hierarchy)**.**   Hierarchy [is] a system of terms related along a single dimension which involves some sort of logical precedence. (Halliday 2002: 42).

Definition 1.2.2 (Taxonomy)**.**   Taxonomy [is] a type of hierarchy with two characteristics:

1. the relation between terms and the immediately following and preceding one is constant

2. the degree is significant and is defined by the place in the order of a term relative to following and preceding terms. (Halliday 2002: 42)

Definition 1.2.3 (Cline)**.**   Cline [is] a hierarchy that instead of being made of a number of discrete terms, is a continuum carrying potentially infinite gradations. (Halliday 2002: 42).

The concept of cline may not necessarily originate in SFL but it is used quite extensively in the domain literature. Next I define and introduce each category of *grammatics* and the related concepts that constitute the theoretical foundation for the Sydney Theory of grammar.

1.2.1   Unit

Language is a patterned activity of meaningful organization. The patterned organization of substance (*graphic* or *phonic*) along a linear progression is called *syntagmatic order* (or simply *order*).

Definition 1.2.4 (Unit)**.**   The unit is a grammatical category that accounts for the stretches that carry grammatical patterns (Halliday 2002: 42). The units carry a fundamental *class* distinction and should be fully identifiable in description (Halliday 2002: 45).

Generalization 1.2.1 (Constituency principles)**.**   The five principles of constituency in lexicogrammar are:

1. There is a scale or rank in the grammar of every language. That of English (typical of many) can be represented as: clause, group/phrase, word, morpheme.

2. Each unit consists of *one or more* units of rank next below.

3. Units of every rank may form complexes.

4. There is potential for rank shift, whereby a unit of one rank my be down-ranked to function in a structure of a unit of its own rank or of a rank below.

5. Under certain circumstances it is possible for one unit to be enclosed within another, not as a constituent but simply in such a way as to split the other into two discrete parts (Halliday & Matthiessen 2013: 9–10).

For example, the down-ranking (Point 4) can be observed in nominal groups that incorporate a relative clause functioning as qualifier. In example 1 *that I got for Christmas* is a relative clause specifying which books are being referred. The unit split (Point 5) can be encountered in the instances of Wh-interrogative clauses containing a preposition at the end which in fact belongs to the Wh-group. In example 2 the prepositional phrase *Who . . . about* is gapped and has an inverted order of constituents.

(1)   I haven't read any books *that I got for Christmas.*

(2)   *Who* are you talking *about*?

(3)   I am talking about George.

The relation between units is that of consistency for which we say that a unit *consists of* other units. The scale on which the units are ranged is the *rank scale*. The rank scale is a levelling system of units supporting unit composition regulating how units are organised at different granularity levels from clause, to groups/phrases to words and the units of a higher rank scale consist of units of the rank next below. Table 1.1 presents a schematic representation of the rank scale and its derived complexes.

Rank scale ↓	Complexing
	Clause complex
Clause	
	Group(/phrase) complex
Group(/phrase)	
	Word complex
Word	
	(Morpheme complex)
(Morpheme)	

Table 1.1 Rank scale of the (English) lexicogrammatical constituency

Generalization 1.2.2 (Rank scale constraints)**.**   The rank relations are constrained as follows:

1. in general elements of clauses are filled by groups,the elements of groups by words and the elements of words by morphemes.

2. downward *rankshift* is allowed i.e. the transfer of a given unit to a lower rank.

3. upward rankshift is not allowed.

4. only whole units can enter into higher units (Halliday 2002: 44).

The Generalization 1.2.2 taken as a whole means that a unit can include, in what it consists of, a unit of rank higher than or equal to itself but not a unit of rank more than one degree lower than itself; and not in any case a part of any unit (Halliday 2002: 42).

Following the rank scale constraints above the concept of embedding can be defined as follows.

Definition 1.2.5 (Embedding)**.**   Embedding is the mechanism whereby a clause or phrase comes to function as a constituent within the structure of a group, which is itself a constituent of a clause. (Halliday & Matthiessen 2013: 242)

Halliday states that embedding is a phenomena that occurs only when a phrase/-group or clause function within the structure of a group which is itself a constituent of a clause (Halliday 1994: 242). The above definition of embedding permits the only for a clause and groups that function as elements of groups which means that a clause cannot fill the elements of another clause (Fawcett 2000: 237).

1.2.2   Structure

Definition 1.2.6 (Structure)**.**   The structure (of a given unit) is the arrangement of *elements* that take places distinguished by the order relationship (Halliday 2002: 46).

Definition 1.2.7 (Element)**.**   Element is defined by the place stated as absolute or relative position in sequence and with reference to the unit next below (Halliday 2002: 47).

We say that a unit is composed of elements located in places and that its internal structure is accounted for elements in terms of functions and places taken by the lower (constituting) units or lexical items. The graphic representation of the unit structure
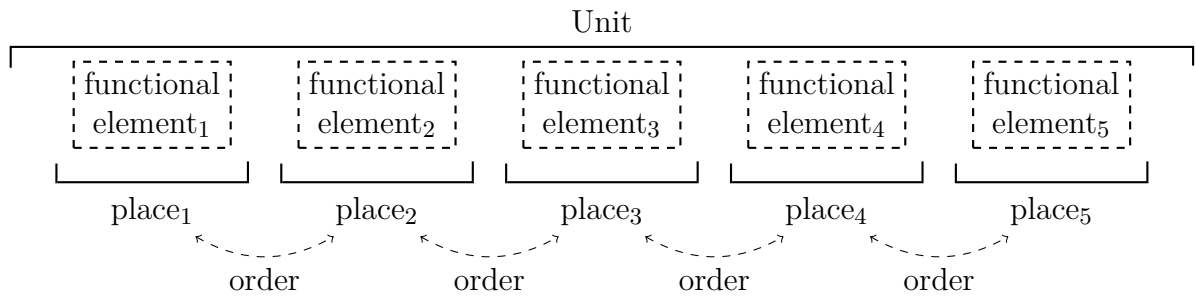
Fig. 1.2 The graphic representation of (unit) structure

is depicted in Figure 1.2. The unit structure is referred in linguistic terminology as *constituency* (whose principles are enumerated in Generalization 1.2.1). In the unit structure, the elements resemble an array of empty slots that are *filled* by other units or lexical items.

For example to account for the English clause structure four elements are needed: *subject*, *predicator*, *complement* and *adjunct*. They yield the distinct symbols, so that S, P, C, A is the inventory of elements. They then can be arranged in various orders falling in particular places, say SPC, SAPA, ASPCC etc. The places of elements are important with respect to the structure of the whole unit but also with respect to the relative ordering between these elements. For example S always fronts P, C is fronted by P unless the clause realises a Wh-interrogative whereas A is quite free and can occur anywhere in the unit structure.

1.2.3 Class

To one place in the structure corresponds one occurrence of the unit next below. This means that there will be a certain grouping of members identified by the functional element they take in the structure. Patterning such groupings leads to emergence of *classes* of units.

In the clause structure example, elements in the unit are occupied by units of lower rank and of a particular class. The relation between the element and the class is mutually determined. In each of these elements is placed a lower rank unit and of an expected class. For instance in the S position can be placed a *noun*, *nominal group*, *pronoun* or another *clause* (that will be a down-ranking situation defined above).

Definition 1.2.8 (Class)**.** The class is that grouping of members of a given unit which is defined by the operation (i.e. functional element) in the structure of the unit next above (Halliday 2002: 49).

Halliday defines class (Definition 1.2.8) as likeness of the same rank *phenomena* to occur together in the structure. He adopts a top-down approach stating that the class of a unit is determined by the *function* (Definition 1.2.13) it plays in the unit above and not by its internal structure of elements. In SFG the structure of each class is well accounted in terms of syntactic variation recognizing six unit classes: *clause*, *nominal*, *verbal*, *adverbial* and *conjunction* groups and *prepositional phrase*. The Sydney unit structure model is briefly summarised in the Appendix **??**.

Halliday identifies the concept of *grammatical metaphor* defined in 1.2.9 and it plays an important role in the SFG as a whole for accounting for the versatility of natural language. It typically found in adult language where one type of process are expressed in the grammar of another.

Definition 1.2.9 (Grammatical metaphor)**.** Grammatical metaphor involves the substitution of one grammatical class or structure for another, often resulting in a more compressed expression.

(4) The fifth day saw them at the summit.

(5) On the fifth day they arrived at the summit.

(6) Guarantee limited to refund of purchase price of goods.

(7) we guarantee only to refund the price for which the goods were purchased.

For example 4 and 6 are instances of grammatical metaphor whereas the 5 and 7 are their non metaphorical counterparts. In Examples 4 and 5 the temporal circumstance of an action expressed through a prepositional phrase becomes the nominal actor of a perception process. Children speech is largely free of such kind of metaphors, in fact this is the main distinctions between the two.

1.2.4 System

As described above, structure is a syntagmatic ordering in language capturing regularities and patterns which can be paraphrased as *what goes together with what*. However in SFG most of the descriptive work is carried not syntagmatically but paradigmatically via *system networks* (Definition 1.2.10) describing *what could go instead of what* (Halliday & Matthiessen 2013: 22). Note that the paradigmatic-syntagmatic axes date back to the works of Saussure (1959 [1915]). Both are important for completing a linguistic description. Here lies one of the main differences between SFL and other approaches which is is taking the paradigmatic path whereas many others take the syntagmatic path to language representing it as an inventory of structures. The structure of course

is a part of language description but it is only a syntagmatic manifestation of the systemic choices and one needs to account for both (Halliday & Matthiessen 2013: 23).

Definition 1.2.10 (System). A system is a set of mutually exclusive set of terms referring to meaning potentials in language and are mutually defining. The system is considered self-contained, closed and complete with the following characteristics:

1. the number of terms is finite,

2. each term is exclusive of all others,

3. if a new term is added to the system it changes the meaning of all the other terms Halliday (2002: 41).

The concept of a system as presented in Definition 1.2.10 has its roots in the works of Saussure (1959 [1915]) and Hjelmslev (1953) and Halliday only cements it in SFL architecture of grammar.

Going back to the notion of class previously defined as a grouping of items identified by functions in the structure, it needs stressed here that class is not a list of formal items but an abstraction from them. By increase in *delicacy* a class is broken into secondary classes.

Definition 1.2.11 (Delicacy). Delicacy is the scale of differentiation or depth of detail whose limit at one end is the primary degree of categories of structure and class and on the other end, theoretically, is the point beyond which no further grammatical relations obtain. Halliday (2002: 58)

We say that a category is refined into more subtle distinctions of subcategories which form a system as define above. Subsequently those distinctions fo subcategories can be further refined in other systems. This relationship between these two systems is one of delicacy where the second one is more delicate than the first one and together they form a *system network*.

The graphical notations introduced by Halliday & Matthiessen (2013) are useful in reading and writing system networks in this thesis. Below is a system network with a simple *entry condition* (Figure 1.3), a *system network grouping* that share the same entry condition (Figure 1.4), a system network with a *disjunctive* and *conjunctive* entry conditions (Figure 1.5 and 1.6).

$$a \rightarrow \begin{array}{l} \text{x} \\ \text{y} \end{array}$$
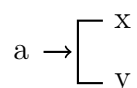
Fig. 1.3 A system with a single entry condition: if a then either x or y

$$a \left\{ \begin{array}{l} \rightarrow \begin{array}{l} \text{x} \\ \text{y} \end{array} \\ \rightarrow \begin{array}{l} \text{p} \\ \text{q} \end{array} \end{array} \right.$$
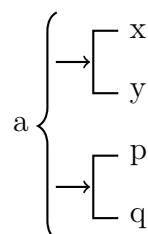
Fig. 1.4 Two systems grouped under the same entry condition: if a then both either x or y and, independently, either p or q

$$\begin{array}{l} a \\ c \end{array} \Big] \rightarrow \begin{array}{l} \text{x} \\ \text{y} \end{array}$$
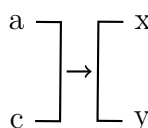
Fig. 1.5 A system network with a disjunctive entry condition: if either a or c (or both), then either x or y

$$\begin{array}{l} a \\ b \end{array} \Big\} \rightarrow \begin{array}{l} \text{x} \\ \text{y} \end{array}$$
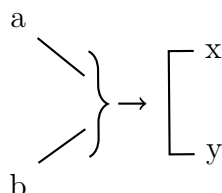
Fig. 1.6 A system with a conjunctive entry condition: if both a and b then, either x or y

It is worth noting that when a piece of language is analysed, it can be approached at various levels of delicacy. We say that delicacy is variable in description, and one may choose to provide coarse grained analysis without going beyond primary grammatical categories or it can dive into fine grained categorial distinctions, still being comprehensive with regards to the rank, *exponence* and grammatical categories.

Definition 1.2.12 (Exponence). Exponence is the scale which relates the categories of theory which are with high degree of abstraction to formal items on its low end. Each exponent can be linked directly to the formal item or by taking successive steps on the exponence scale and changing rank where necessary. Halliday (2002: 57)

And in relation to the previous section, the class stand in the relation of exponence to an element of primary structure of the unit next above. This breakdown gives a system of classes that constitute choices implied by the nature of the class (Halliday 2002: 41).

1.2.5 Functions and metafunction

Above, when talking about structure, I described a unit as being composed of elements accounted in terms of *functions* and places taken by the lower (constituting) units or lexical items.

Definition 1.2.13 (Function). The functional categories or functions provide an interpretation of grammatical structure in terms of the overall meaning potential of the language. (Halliday & Matthiessen 2013: 76).

Most constituents of clause structure, however, have more than one function, which is called a *conflation of elements*. For example in the sentence "Bill gave Dolly a rose", "Bill" is the Actor doing the act of giving but also the Subject of the sentence. So we say that Actor and Subject functions are conflated in the constituent "Bill". This is the concept of *metafunction* or *strand of meaning* comes into the picture. The Subject function is said to belong to the *interpersonal metafunction* while the Actor function belongs to the *experiential metafunction*.

Halliday identifies three fundamental dimensions of structure in the clause, each meaning: *experiential, interpersonal* and *textual*. He refers to them as *metafunctions* and they account for the functions that language units take on in communication. Table 1.2 presents the metafunctions and their reflexes in grammar as proposed by Halliday & Matthiessen (2013: 85).

Metafunction	Definition(kind of meaning)	Corresponding status in clause	Favored type of structure
experiential	construing a model of experience	clause as representation	segmental (based on constituency)
interpresonal	enacting social relationship	clause as exchange	prosodic
textual	creating relevance to context	clause as message	culminative
logical	constructing logical relations	complexes (taxis & logico-semantic type)	iterative

Table 1.2 Metafunctions and their reflexes in the grammar

Across the rank scale, with respect to structure and metafunctions, Halliday formulates the general principle of *exhaustiveness* (Generalization 1.2.3) saying that clause constituents have at least one and may have multiple functions in different strands of meaning; however this does not mean that it must have a function in each of them. For example interpersonal Adjuncts such as "perhaps" ot textual Adjuncts such as "however" play no role in the clause as representation.

Generalization 1.2.3 (Exhaustiveness principle)**.** Everything in the wording has some function at every rank but not everything has a function in every dimension of structure (Halliday 2002; Halliday & Matthiessen 2013).

This principle implicitly relates to the property of language meaning that there is nothing meaningless and thus every piece of language must be explained and accounted for in the lexicogrammar. Also this principle implies that each metafunction has its own structure or that text is analysed through a multi structural approach.

At the very top of the rank scale, clauses form complex structures. Halliday employs systematically the concepts of *taxis* and *logico-semantic relations* to account for inter-clausal relations.

Definition 1.2.14 (Taxis)**.** *Taxis* represents the degree of interdependency between units systematically arranged in a linear sequence where *parataxis* means equal and *hypotaxis* means unequal status of units forming a *nexus* or a *unit complex* together.

The concept of taxis which is very useful at describing unit relations not only at the group and clause ranks but all the way down to smallest linguistic unit such as

morphemes and phonemes. I will also refer to it when describing the Cardiff theory of grammar and also briefly in the discussion of dependency relations in Section **??**.

The elements of logical paratactic structure are notated left to right with numbers (1   2) while those of hypotactic structure with Greek letters (β α) right to left. The tactic relations can be of two types: that of expansion which relates phenomena of the same order of experience and that of projection which relates phenomena of one order of experience (usually saying or thinking) to an order of experience higher (what is said or thought). Projection can be of two types: *idea* (' single quote) and *locution* (" double quotes).

Expansion is further divided into three: *elaborating* ($=$ equals), *extending* ($+$ is added to) and *enhancing* ($\times$ is multiplied by). Elaboration is a way to restate the same thing, exemplify, comment or specify in detail. Extending is the way to add new element give an exception or offer an alternative. And finally enhancing is the way to qualify something with some circumstantial feature of time, place, cause, intensity or condition.

1.2.6   Lexis and lexicogrammar

In SFL the terms *word* and *lexical item* are not really synonymous. They are related but they refer to different things. The term *word* is reserved (in early Halliday) for the grammatical unit of the lowest rank whose *exponents* are lexical items.

Definition 1.2.15 (Lexical Item)**.**   In English, a lexical item may be a *morpheme*, *word* (in traditional sense) or *group (of words)* and it is assigned to no rank (Halliday 2002: 60).

Examples of lexical items are the following: "'s" (the possessive morpheme), "house", "walk", "on" (words in traditional sense) and "in front of", "according to", "ask around", "add up to", "break down" (multi word prepositions and phrasal verbs).

If some theories treat grammar and lexis as discrete phenomena, Halliday brings them together as opposite poles of the same cline. He refers to this merge as *lexicogrammar* where they are paradigmatically related through delicacy relation. Hasan (2014), explores the feasibility of what would it mean to turn the "whole linguistic form into grammar". This then implies a assumption that lexis is not form and that its relation to semantics is unique which in turn is challenging the problems of polysemy.

1.3 The Cardiff theory of grammar

As presented in the introduction and explained by Bateman (2008), the accounts along the syntagmatic axis had gone missing in the Sydney grammar leaving unresolved how to best represent the structure of language at the level of form. This section present the theory of systemic functional grammar as conceived by Robin Fawcett at the University of Cardiff. His book "A theory of syntax for Systemic Functional Linguistics" (Fawcett 2000) presented a proposal for a *unified syntactic model* for SFL that contrasts several aspects of Hallidayan grammar but share the same set of fundamental assumptions about the language; it is an extension and a simplification in a way.

Fawcett questions the status of multiple structures in the theory and whether they can finally be integrated into a simpler sole representation. A big difference to Hallidayan theory is renouncing the concept of rank scale which has an impact on the whole theory. Another is the bottom-up approach to unit definition as opposed to top-down one advocated by Halliday. These two and a few other differences have important implications for the overall theory of grammar and consequently for the grammar itself. As a consequence, to accommodate the lack of rank-scale, Fawcett adapts the definitions of the fundamental concepts and changes his choice of words (for example "class" and "unit" turn into "class of unit" treated as one concept rather than two distinct ones).

Fawcett (2000) proposes three fundamental categories in the theory of grammar: *class of unit*, *element of structure* and *item*. Constituency is a relation accounting for the prominent compositional dimension of language. However a unit does not function directly as a constituent of another unit but via a specialised relation which Fawcett breaks down into three sub-relations: *componence*, *filling* and *exponence*. Informally it is said that a unit is composed of elements which are either filled by another unit or expounded by an item. He also proposes three secondary relations of *coordination*, *embedding* and *reiteration* to account for a more complete range of syntactic phenomena.

1.3.1 Class of units

Fawcett's theory of language assumes a model with two levels of *meaning* and *form* corresponding to *semantic units* and *syntactic units* which are mutually determined (which is the case for any sign in a Saussurean approach to language).

Definition 1.3.1 (Class of Unit)**.** The class of unit [...] expresses a specific array of meanings that are associated with each one of the major classes of entity in semantics

[...and] are to be identified by the elements of their internal structure (Fawcett 2000: 195).

For English Fawcett proposes four main kinds of semantic entities: situations, things, qualities (of both situations and things) and quantities. Each of these semantic units corresponds to five major classes of syntactic units: *clause*, *nominal group*, *prepositional group*, *quality group* and *quantity group*. In addition he recognises two more minor classes i.e. the *genitive cluster* and the *proper name cluster* (Fawcett 2000: 193–194).

Fawcett's classification is based on the idea that the syntactic and semantic units are mutually determined and supported by grammatical patterns. However those patterns lie beyond the syntactic variations of the grammar and so blend into lexical semantics.

In Sydney theory the class is determined by the function it plays in the unit above. By contrast, in Cardiff theory, the class of unit is determined based on its internal structure i.e. by its *elements of structure* (and not by the function it plays in the parent unit).

1.3.2   Element of structure

The terms *element* and *structure* have roughly the same meaning as defined in Sydney theory of grammar (defined in Section 1.2) but with two additional stipulations presented below.

Definition 1.3.2 (Element of Structure).   Elements of structure are immediate components of classes of units and are defined in terms of their *function* in expressing meaning and not in terms of their absolute or relative position in the unit. (Fawcett 2000: 213–214).

The definition above leads as a consequences to two important properties of elements formulated as follows.

Generalization 1.3.1 (Element functional uniqueness).   Every element in a given class of unit serves a function in that unit different from the function of the sibling elements (Fawcett 2000: 214).

Even if for example, different types of *modifiers* in English nominal group seem to have very slight differences in functions, they are still there.

Generalization 1.3.2 (Element descriptive uniqueness)**.** Every element in every class of unit will be different from every element in every other class of unit (Fawcett 2000: 214).

Thus the terms of modifier and head shall not be used for more than one class of unit. In English grammar the head and modifier are used for nominal group only. And in other groups the elements of structure may seem similar to modifier and head, they still receive different names such as *apex* and *temperer* in the quality group.

The elements (of structure) are functional slots which define the internal structure of a unit but still they are *located* in *places*. One more category that intervenes between element and unit is the concept of *place* which become essential for the generative versions of grammar.

There are two ways to approach place definition. The first is to treat places as positions of elements relative to each other (usually previous). This leads to the need for an *anchor* or a *pivotal element* which may not always be present/realised.

The second is to treat places as a linear sequence of locations at which elements may be located, identified by numbers "place 1", "place 2" etc. This place assignment approach is absolute within the unit structure and makes elements independent of each other. This approach has been used in COMMUNAL (Fawcett 1990) and PENMAN (Mann 1983b) projects.

1.3.3   Item

Definition 1.3.3 (Item)**.** The item is a lexical manifestation of meaning outside syntax corresponding to both words (in the traditional sense), morphemes and either intonation or punctuation (depending whether the text is spoken or written). (Fawcett 2000: 226–232).

Items correspond to the leaves of syntactic trees and constitute the raw *phonetic* or *graphic* manifestation of language. The collection of items of a language is generally referred to as *lexis.*

Since items and units are of different natures, the relationship between an element and a (lexical) item must be different from that to a unit. We say that items *expound* elements and not that they *fill* elements as units do.

Definition 1.3.4 (Exponence (restricted))**.** Exponence is the relation by which an element of structure is realised by a (lexical) item (Fawcett 2000: 254).

If in Sydney model exponence (Definition 1.2.12) is a relation that links abstract grammatical categorise to the data In Cardiff model it has a restricted meaning referring to relation between items and elements only.

1.3.4   Componence and obscured dependency

Definition 1.3.5 (Componence).   Componence is the part-whole relationship between a unit and the elements it is composed of (Fawcett 2000: 244).

Note that componence is not a relationship between a unit and its places; the latter, as discussed in Section 1.3.2, simply locationally relate elements of a unit to each other.

Componence intuitively implies a part-whole constituency relationship between the unit and its elements. But this is not the only view. Another perspective is the concept of *dependency* (which I will address in Chapter **??**) or strictly speaking the *sister* or *sibling dependency* (not parent-daughter). It is suitable for describing relations between elements of structure within a unit.

(8)   the man with a stick

For example the componence of nominal group in Example 8 is (*dd h q*) which are symbols for (*determiner head qualifier*). The same can be expressed in terms of sibling dependency relations depicted in Figure 1.7. The relations from *stick* to *with a* are not depicted because they belong in description of prepositional group *with a stick*.
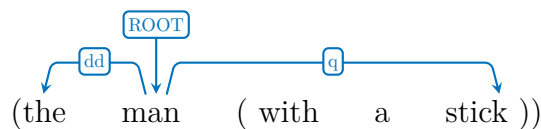
Fig. 1.7 Sibling dependency representation for "the man with a stick"

In both SFL theories, the sister dependency relations is considered a by-product or second order concept that can be deduced from the constituency structure thus unnecessary in the grammar model. I will come back to this point because current work relies on this dual view on elements of structure and relation to the whole unit.

The (supposed) dependency relation between a modifier and the head, in the framework of SFG is, not a direct one. Simply assume that what modifier modifies is the head. Here, however, the general function of the modifiers is to contribute to the meaning of the whole unit which is anchored by the head.

In the nominal group from Example 8, the *determiner* and *qualifier* are modifiers that contributes to the description of the referent stated by the *head*. So the head

realises one type of meaning that relates the *referent* while modifier realises another one. Both of them describe the referent via different kinds of meaning, therefore, according to Fawcett, they are related indirectly to each other because the modifier does not modify the head but the referent denoted by the head From this point of view, whether the element is dependent on a sibling element such as the head or on the parent unit is beside the point because in syntax we can observe its realization in system networks (Fawcett 2000: 216–217). Next I move towards last concept in Cardiff model, *filling*, which is a relation between the elements of structure and the units below.

1.3.5 Filling and the role of probabilities

Definition 1.3.6 (Filling)**.** Filling is the probabilistic relationship between a element and the unit lower in the tree that operates at that element (Fawcett 2000: 238, 251).

Fawcett replaces the rank scale with the concept of *filling probabilities*. The probabilistic predictions are made in terms of filling relationship between a unit and an element of structure in a higher unit in the tree rather than being a relationship between units of different ranks. This moves focus away from the fact that a unit is for example a group, and towards what group class it is.

In this line of thought, some elements of a clause are frequently filled by groups, but some other elements are rather *expounded* by items. The frequency varies greatly and is an important factor for predicting or recognizing either the unit class or the element type in the filling relationship.

Filling may add a *single unit* to the element of structure or it can introduce *multiple coordinated units*. Coordination (Example 9) is usually marked by an overt *Linker* such as *and*, *or*, *but*, etc. and sometimes is enforced by another linker that introduces the first unit such as *both*.

Definition 1.3.7 (Coordination)**.** Coordination is the relation between units that fill the same element of structure (Fawcett 2000: 263).

(9)   she is (friendly, nice and polite)

(10)   she is (very very) nice!

Coordination is through by Fawcett as being not between syntactic units but between mental referents. It always introduces more than one unit which are syntactically and semantically in similar (somehow) resulting in a *syntactic parallelism* which often leads to *ellipsis*.

Definition 1.3.8 (Reiteration)**.** Reiteration is the relation between successive occurrences of the same item expounding the same element of structure (Fawcett 2000: 271).

Reiteration (Example 10) often is used to create the effect of emphasis. Like coordination, reiteration is a relation between entities that fill the same element of the unit structure which is problematic in my opinion and I further discuss it in Section 1.4.6.

Filling also makes possible the embedding relation which Fawcett treats as a general principle in contrast to more specific Definition 1.2.5 from Sydney model.

Definition 1.3.9 (Embedding (generic))**.** Embedding is the relation that occurs when a unit fills (directly or indirectly) an element of the same class of units; that is when a unit of the same class occurs (immediately) above it in the tree structure (Fawcett 2000: 264).

(11)   (To become an opera singer) takes years of training.

(12)   The girl (whom he is talking to) is an opera singer.

In Example 11 we can see an occurrence of direct embedding where an infinite clause acts as the subject of another clause. In Example 12 the embedding is indirect as the relative clause is part of the nominal group which functions as the subject in the parent clause. In both cases we say that a lower clause is embedded (directly or indirectly) in higher or parent clause. I will further discuss this in the context of rank-scale concept in Section 1.4.1.

A situation converse to reiteration and coordination where a element is filled by more than one unit, is known as *conflation* where a unit can take more than one function within another.

Definition 1.3.10 (Conflation)**.** Conflation is the relationship between two elements that are filled by the same unit having the meaning of "immediately after and fused with" and function as one element (Fawcett 2000: 249–250).

Conflation is useful in expressing multi-faceted nature of language when for example syntactic and semantic elements/functions are realised by the same unit. For example the Subject "the girl whom he is talking to" is also a *Carrier* while the Complement "an opera singer" is also an *Attribute*. Also conflation relations frequently occur between syntactic elements as well such as for example the *Main Verb* and *Operator* or *Operator* and *Auxiliary Verb*.

Cardiff Grammar in case of both coordination and embedding relations deals without *hypotaxis* and *parataxis* relations described in Sydney Grammar.

Note also that filling and componence are two complementary relations that occur in the syntactic tree down to the level when the analysis moves out of abstract syntactic categories to more concrete category of items via the relationship of exponence.

1.4 Critical discussion on both theories: consequences and decisions for parsing

The two sections above cover the definitions and fundamental concepts from each of the two systemic functional theories of grammar. The work in this thesis uses a mix of concepts from both theories and this section discusses in detail what is being adopted and why attempting a rather pragmatic reconciliation for the purposes of achieving a parsing system than a theoretical debate. Next I draw parallels and highlight correspondences between the Sydney and Cardiff theories of grammar and where alter and present my position on the matter.

1.4.1 Relaxing the rank scale

The *rank scale* proposed by Halliday (2002) became over time a highly controversial concept in linguistics. The discussion whether it is suitable for grammatical description or not still continues. The historic development of this debate is documented in some detail (Fawcett 2000: 309–338).

In this section I present a few cases that highlighting when the rank scale as defined by Sydney is too rigid. As a consequence for the purpose of thesis I drop the *rank scale constraints* as enunciated in Generalization 1.2.2. Also the *rankshift* operation, exceptionally employed to accommodate special cases, is overridden by a broad definition of *embedding* operation (Definition 1.3.9) treated as naturally occurring phenomena in language at all ranks. I do not entirely dismiss the concept of rank scale as proposed in Cardiff school as I still find it useful in classification of units.

(13)   some very small wooden ones

Consider the nominal group 13. Here the modifying element, the Epithet "very small", is not a single word but a group (Halliday & Matthiessen 2013: 390–396). As the rank scale constraints mentioned above state that the group elements need to be filled by words. To account for this phenomena, Halliday, introduces a *substructure* of

modifiers and heads leading to a logical structure analysis as the one in Table 1.3. In such a structure the modifier is further broken down into a Sub-Head and Sub-Modifiers.

some	*very*	*small*	*wooden*	*ones*
Modifier				Head
δ	γ		β	α
	Sub-Modifier	Sub-Head		
	$\gamma\beta$	$\gamma\alpha$		

Table 1.3 Sydney logical structure analysis of Example 13

The corresponding experiential structure analysis is provided in the Table 1.4 Halliday & Matthiessen (2013: 391). Accordingly, the Epithet "very small" is composed of a quality adjective "small" and an enhancer modifier "very".

some	*very*	*small*	*wooden*	ones
Deictic	Epithet		Classifier	Thing
	Sub-Modifier	Sub-Head		

Table 1.4 Sydney experiential analysis of Example 13

As you can see the elements are further broken down into sub-elements composing in a way a structure of their own. This is possible because of the poly-structural and multi functional approach to text analysis which in this case leads to a complex structure of a nominal group. This kind of intricate cases can be simplified through the permission that elements of a group to be filled by other groups or expounded by words. This way, instead of having a sub-modifier construction simply consider that the Epithet is filled by an adjectival or nominal group which in turn has its own structure. Please note that I mention adjectival or nominal group because in Sydney grammar the adjectival group is considered as a nominal group with covert Thing where the Epithet acts as Head; this however is a discussion beyond the point I make here.

The same example analysed with Cardiff grammar would look like in Table 1.5. It follows precisely the above suggestion of filling the Epithet with another unit, in this case a Quality Group which in turn has its own internal structure.

some	*very*	*small*	*wooden*	*ones*
Quantifying Determiner	Modifier		Modifier	Head
	Quality Group			
	Degree Tamperer	Apex		

Table 1.5 Cardiff analysis of Example 13

(14)   Indians had originally planned to present the document to President Fernando Henrique Cardoso.

Indians	had	originally	planned to present	the document	to President Fernando Henrique Cardoso
Mood		Residue			
Subject	Finite	Adjunct	Predicator	Complement	Adjunct
nominal group		adverbial group / verbal group		nominal group	prepositional phrase

Table 1.6 Sydney grammar Mood analysis of Example 14

Another case that deems the rank scale constraints too strict for the present work is in the case of Finite element in the Clause. Consider example 14 where the Finite and Predicator elements are filled by a single unit which is the verbal group which is against the constituency principles which restricts the composition relation to engage only with whole units.

Alternatively, if the unit filling the Finite element is considered separate from the verbal group filling the Predicator then it is always a single word, a modal verb, and never a verbal group. This again is a breach in the rank scale constraints which postulates that a unit may be composed of units of equal rank or a rank higher and cannot be composed of units that are more than one rank lower thus it is not permitted to have clause elements expounded by words directly.

The two cases above I use to demonstrate how the ranks scale construct as defined by Sydney grammar is too rigid and thus unsuitable for the current work. I drop the constituency constraints hence allowing the flexibility for elements to be filled by other units or, in other words, allow unit *embedding*. This approach removes the need of sub-structures in the unit elements reducing thus the structural complexity as seen in Table 1.5.

The weakening of constituency constraints makes embedding a normal (broadly defined in Definition 1.3.9) rather than an exceptional phenomena (strictly defined in Definition 1.2.5).

An approach to describe units outside the rank-scale was suggested by Fawcett (2000) and Butler (1985). Fawcett proposes replacing it with the filling probabilities to guide the unit composition simply mapping elements to a set of legal unit classes that may fill it. Units are carriers of a grammatical pattern, they can be described in terms of their internal structure instead of their potential for operation in the unit above. Nonetheless I do not abandon the rank scale completely and I use it as the top level classifier of grammatical units (see Figure 1.10) falling in line with more traditional syntactic classes.

1.4.2   Approach to structure formation

The *unit* and *structure* are two out of the four fundamental categories in the systemic theories of grammar. The Sydney and Cardiff theories vary in their perspectives on *unit* and *structure* influencing how units are defined and identified.

For Halliday, the *structure* (Definition 1.2.6) characterises each unit as a carrier of a pattern of a particular order of *elements*. The order is not necessarily a linear realisation sequence but a theoretical relation of relative or absolute placement. This perspective has been demonstrated to be useful in generation where unit placement emerges out of the realisation process.

The Cardiff School takes a bottom up approach and defines class in terms of its internal structure describing a relative or absolute order of elements. This sort of syntagmatic account is precisely what is deemed useful in parsing and is the one adopted in this thesis. In this work, as motivated in the Introduction, generation of the constituency structure is derived from the Stanford dependency parse trees. As it consists of words and relations between them the intuitive approach to form groups, clauses and complexes is by working them out bottom up. The method is to let the unit class emerge from recognition of constituent word classes and dependency relations between or sequence of already formed lower units. The exact mechanism how it is done I explain in Chapter **??**. What is important to note here is the bottom up approach which is in line with Cardiff way of defining unit classes in contrast to top down approach of Sydney school.

1.4.3   Relation typology in the system networks

As system is expanded in delicacy to forms a systemic network of choices, choice of a feature in one system becomes the *entry condition* for choices in more delicate systems below. Halliday states that the relation on the systemic cline of delicacy is essentially of *sub-categorisation* (see Definition 1.2.11). In this subsection argue for occurrence of multiple kinds of inter-systemic relations. I also call them *activations relations* because in the traversal process from less to more delicate systems, when choices are made in the former then choice making is enabled or *activated* in the latter.

Next I present a distinction between two activation relations: the *sub-categorisation* and *choice enabling.* that are of interest in the present thesis but this is by no means exhaustive distinction and more work is needed here.

Lets take as example the polarity system represented in figure 1.8. It contains two choices either positive or negative. An increase in delicacy can be seen as a taxonomic

"is a" relationship between features of higher systems and lower systems like in the case of POLARITY TYPE and NEGATIVE TYPE in figure 1.8 and in fact for the rest of the network. As a side note, the delicacy in a system network is akin to sub-classification relation, which was originally the intended one and the predominant one. In practice, however, a few kinds of abstraction relations can be encountered (e.g abstraction as information reduction, as approximation, as idealisation etc.) extensively treated by Saitta & Zucker (2013). This discussion however beyond the scope of the current work.
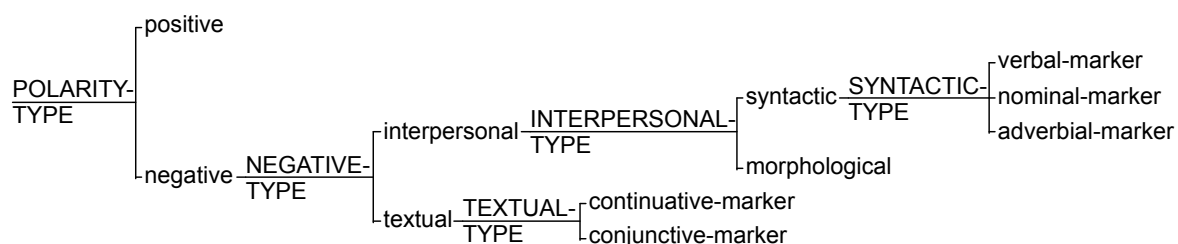
Fig. 1.8 System network of POLARITY

But the activation relation among systems in the cline of delicacy is not always taxonomic. Another relation is "enables selection of" without any sub-categorisation implied. As an example see the FINITENESS system in Figure 1.9 where in case that the finite option is selected then what this choice enables is not subtypes of finite but merely other systems that become available i.e. DEIXIS and INDICATIVE TYPE. The latter is there because selection of finite implies also selection of indicative feature in a sibling of FINITNES system, MOOD-TYPE comprised of options indicative and imperative.
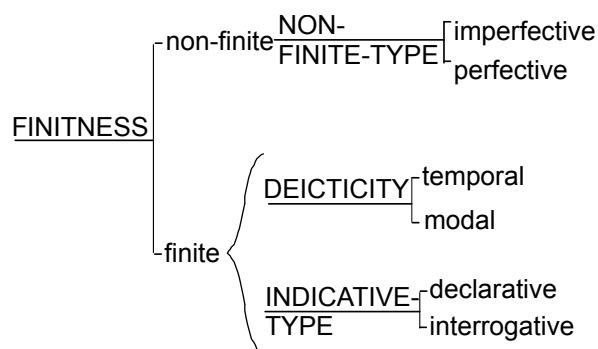
Fig. 1.9 A fraction of the finiteness system where increase of delicacy is not a "is a" relation

The distinction in the systemic relations is incorporated into the technical data structure definitions and traversal algorithms proposed in the Chapter **??**.

1.4.4 Unit classes

In SFL at large there is the consensus that linguistic forms and meanings are intertwined and mutually determined just like for any sign in a Saussurean approach to language. Both Halliday (quote below) and Fawcett (Definition 1.3.1) adopt this position.

> . . . something that is distinctly non-arbitrary [in language] is the way different kinds of meaning in language are expressed by different kinds of grammatical structure, as appears when linguistic structure is interpreted in functional terms (Halliday 2003a).

When it comes to establishing the lexicogrammatical classes the two schools diverge. Halliday adopts the traditional grammar *word classes* or *parts of speech*: noun, verb, adjective etc. He then derives a set of groups (e.g. nominal group, verbal group, adverbial group etc.) that share properties of the word classes. In fact the class, in Halliday's words, "indicates the in a general way its potential range of grammatical functions" (Halliday & Matthiessen 2013: 76). For example the nominal group is a formation that functions as a noun may do and expresses same kind of meaning.

Following the idea that major semantic classes of entities (situations, things, qualities and quantities) correspond to the major syntactic units, Fawcett decided to mirror them into the lexicogrammar. This lead to a semantically based classification of syntactic units: clause, nominal group, prepositional group, quality group and quantity group (Fawcett 2000: 193–194) along with a set of minor classes such as genitive and proper name clusters. This is, in a way, a tight coupling of the grammatical units with an ontology which may be subject to change in the future. The converse may also be stated that the traditional part of speech are disconnected from the semantics in the sense that there is no one to one correspondence (as Fawcett attempts) but rather complex set of mappings. Establishing the exact interface of syntax and semantics is a hot ongoing theoretical exploration across the entire linguistic discipline a difficult task in practice. This discussion however is beyond the scope here.

In the current work I side with the Sydney classification of syntactic units that is close in line with traditional syntactic classifications (Quirk et al. 1985). I adopt the clause as a unit plus the four group classes of the Sidney grammar depicted in Figure 1.10.

(a) The group classes
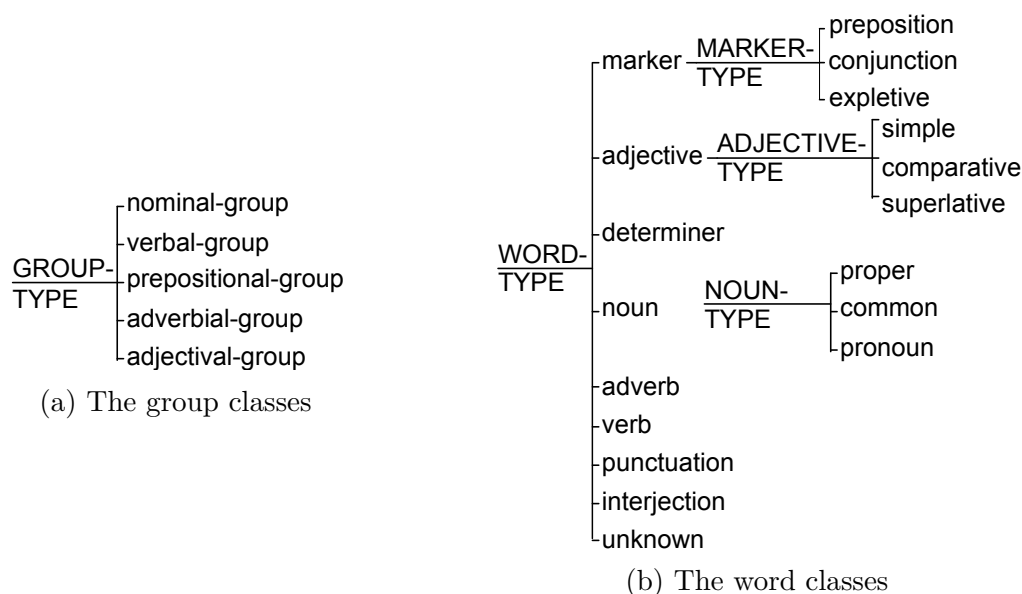
(b) The word classes

Fig. 1.10 The group and word classes

The word classes or part of speech tags that I adopt here are the ones employed to annotate Penn Treebank corpora called the Penn tag set (Marcus et al. 1993) which, like Sidney unit classes, are also in line with the traditional grammar. This tag set has become a widely accepted standard in mainstream computational linguistics and there are multiple implementation of the part of speech taggers. The Stanford Parser which plays an important role in the software implementation of this thesis and is described in the Chapter **??**, employs precisely the Penn tag set.

The Penn tag set was developed to annotate the Penn Treebank corpora (Marcus et al. 1993). It is a large, richly articulated tag set that provides distinct codings for classes of words that have distinct grammatical behaviour.

The Penn tag set is based on the Brown Corpus tag set (Kucera & Francis 1968) but differs in several ways. First, the authors reduced the lexical and syntactic redundancy. In the Brown corpus there are many unique tags to a lexical item. In the Penn tag set the intention is to reduce this phenomenon to a minimum. Also distinctions that are recoverable from lexical variation of the same word such as verb or adjective forms or distinctions recoverable from syntactic structure are reduced to a single tag.

Second the Penn Corpus takes into consideration the syntactic context. Thus the Penn tags, to a degree, encodes syntactic functions when possible. For example, *one* is tagged as NN (singular common noun) when it is the head of the noun phrases rather than CD (cardinal number).

Third, Penn POS set allows multiple tags per word, meaning that the annotators may be unsure of which one to choose in certain cases. There are 36 main POS and 12 other tags in the Penn tag set. A detailed description of the schema, the design principles and annotation guidelines are described in (Santorini 1990). Figure 1.10 depicts a classification summarising the Penn tag set.

1.4.5   Syntactic and semantic heads

In SFG the heads may be motivated by semantic or syntactic criteria (simply called here semantic or syntactic heads). In most cases they coincide but there are exceptions when they differ or even diverge. This topic is especially important in the discussions of the **nominal group** structure (continued in Section 1.5.3) on which Halliday & Matthiessen (2013) offers a thorough examination and Fawcett (2000) provides a more generic perspective.

In this discussion I show few examples when the syntactic and semantic heads diverge and argue my position on the group formation on two points. First, the class of the Head (in Sydney school) or pivotal element (in Cardiff school) is not always raised to establish the group class but the whole underlying structure determines the group class. Second, that syntactically motivated heads are easy to establish because they are based solely on the formal grounds whereas the semantic heads require an evaluation at the level of entire group, once one is established, employing additional lexical semantic resources. This can be a two step process but in the current implementation only the group structure on syntactic grounds is provided.

As mentioned before in Section 1.2.5, Sydney grammar fulfils the exhaustiveness principle (Generalization 1.2.3) through multiple parallel structures while Cardiff grammar through a single syntactic structure resembling a mixture of the former.

Let's briefly return to the Example 13 analysed with Sydney grammar in Table 1.3 and 1.4 that reflect the nominal group logical and experiential structures Halliday & Matthiessen (2013: 391). When the Head (called here the *syntactic head* of the nominal group) coincides with the Thing (called here the *semantic head*) we say that they are conflated (Definition 1.3.10) and examples as this one may lead to the assumption that the Head, which is motivated by syntactic criteria, is also always the Thing which is motivated by the semantic criteria, but this is not so.

The logical structure is a Head-Modifier structure and "represents the generalised logical-semantic relations that are encoded in the natural language" (Halliday & Matthiessen 2013: 388). The experiential structure of the nominal group as a whole has the function of specifying the class of things, through the Thing element, and some

category of membership in this class, through the rest of the elements. In the nominal group there is always a Head but the Thing may be missing and so the Head element is conflated with either Epithet, Numerative, Classifier or Deictic instead.

(15)   (Have) a cup of tea.

(16)   The old shall pass first.

(17)   I'll give you three.

Consider Example 15 analysed with Sydney and Cardiff grammars in Table 1.7. In the Sydney Grammar the semantic and the syntactic heads differ. In the experiential analysis the semantic head is "tea" which functions as Thing, while in the logical analysis the syntactic head is "cup" which functions as Head. Cardiff Grammar does not offer multi-structural analysis and there is no Head/Thing distinction. The functional elements are already established based on semantic criteria and this is further discussed in Section 1.5.3.

		a	*cup*	*of*	tea
Sidney Grammar	experiential	Numerative			Thing
	interpersonal	Pre-Modifier	Head	Post-Modifier	
Cardiff Grammar		Quantifying Determiner		Selector	Head

Table 1.7 Analysis of Example 15 with Sydney and Cardiff grammars: diverging semantic and syntactic heads.

In the nominal group "The old" which is Subject in Example 16, the Head is adjective "old" and not a noun as would normally be expected. The noun modified by the adjective "old", also the *pivotal element* of the group defined in Section 1.3.2, is left covert and it should consequently be recoverable anaphorically or cataphorically from the context. We can insert a generic noun "one" to form a canonical noun group "the old one". In such cases when the pivotal noun is missing, the logical Head is conflated with other element in this case the Epithet. The group class is not raised from the word class to quality group but is identified by internal structure of the whole group and in this case the presence of determiner signals a nominal class. Similarly, in Example 17, "three" in Sydney grammar is a nominal group where the Thing is missing and the Head has shifted left towards the Numeral. With examples as these ones, Fawcett argues that none of the constituting elements of the unit is mandatorily realised, even the so called pivotal element which is the group defining element. In Chapter **??** is provided an in depth description of the recovering mechanisms for covert nominal elements at the level of the clause.

In this work I adopt the principles for establishing the logical structure of Sydney Grammar. It resonates closely with the traditional "semantically blinded" grammars because and it always provides a Head element even if it differs from the syntactically motivated pivotal element in Cardiff Grammar. Moreover these logical Heads correspond to dependency heads established in the Stanford dependency parse. Chapter **??** provides the grounds for cross-theoretical mappings and the empirical evaluation in Chapter **??** validates it.

In language is not unusual to have nominal groups with the Thing missing or elliptic clauses with the Main verb missing therefore no rigid correspondence can be established between the logical Head and unit class. In this thesis the structure creation is performed in two steps first establishing the group boundaries and the unambiguous unit elements through a top down perspective (that is Sydney approach to unit creation) and second for each established group evaluate the internal structure in order to establish the group class (that is Cardiff approach to group formation). This process is detailed in the Chapter **??**.

The evaluation in the second step, besides finalising the syntactically motivated unit structure, can as well assign semantically motivated unit structure. This part however is left out in the current thesis for the groups and only the clauses receive semantic role labels and process types described in Chapter **??**.

1.4.6   Coordination as unit complexing

In Sydney Grammar unit complexes fill an important part of the grammar along with the *taxis relations* (Definition 1.2.14) which express the interdependency relations in unit complexes. *Parataxis* relations bind units of equal status while the *Hypotaxis* ones bind the dominant and the dependant units. Fawcett bypasses the taxis relations replacing them with coordination and embedding (Fawcett 2000: 271) leading to abandonment of unit complexing entirely. While embedding elegantly accounts for the depth and complexity of syntax, this approach to coordination is problematic.

Hereafter I discuss the utility and even necessity of keeping unit complexes in parsing. In particular I address the treatment of group and clause *coordination* but the same principle applies to fixed idiomatic structures such as *comparatives*, *conditionals* or *appositions*.

Treatment of the coordination phenomena is a challenge not only for SFL but for other linguistic theories as well. Sydney Grammar approaches it through unit complexing and taxis relations while Cardiff Grammar treats this phenomena as multiple distinct units filling or expounding the same element.

Table 1.8 illustrates an example analysis where the Complement is filled by a homogeneous nominal group complex held together through *paratactic extension* where the first element is a nominal group and the second is nominal group together with the conjunction which is not part of the experiential structure but remains accounted only in the logical structure of the nexus.

Ike	washed	his	shirt	and	his	jeans
Subject	Predicate/Finite	Complement				
		1		+2		
		Deictic	Thing		Deictic	Thing

Table 1.8 Clause with nominal group complex

In Table 1.9 the Epithet is filled by a nexus of paratactic extension. The first element of the nexus is the word "immediate" and the second element is the set of words "and not so far distant". The "not so far distant" is an adverbial group with a logical structure of sub-modifiers already discussed in Section 1.4.1 and the conjunction "and" is left implicitly part of the logical structure of the nexus creating a gap in the structure that is addressed in this discussion. Also note that, in Sydney Grammar the coordination is accounted as unit complex ensuring that only one unit fills an element of the parent, in contrast, as we will see below, to Cardiff Grammar.

the	immediate	and	not	so	far	distant	future
Modifier							Head
γ	β						α
Deictic	Epithet						Thing
	1	+2					
			Sub-Modifier			Sub-Head	
			δ	γ	β	α	

Table 1.9 Nominal group with word complex from (Halliday & Matthiessen 2013: 564)

In Table 1.10 is presented an example of analysis with Cardiff Grammar. The Complement is filled by two *sibling* nominal groups "his shirt" and "and his jeans" that are both of them fill the same element in accordance to Definition refdef:coordination. The conjunction "and" is accounted directly as part of the nominal group structure.

Opinions are divided (between Sydney and Cardiff schools) whether to invite the notion of a complex unit to handle coordination or not. If we side with Cardiff grammar and dismiss the unit complex then we allow an element to be filled by more than

Ike	*washed*	*his*	*shirt*	*and*	*his*	*jeans*
Subject	Main Verb	Complement				
		Deictic Determiner	Head	&	Deictic Determiner	Head

Table 1.10 Coordination analysis in Cardiff Grammar

one units. And this is a problem because if we do not account unit elements each in a unique place within the unit structure then we loose the capacity to order them. Therefore in this thesis I adopt the Sydney definition of structure (Definition 1.2.6) that constraints each element into a single place that is filled by another unit. Therefore the conjunction must be a nexus acting as a single unit filling a single element.

I argue for adoption of such unit type in order to ensure that maximum one unit can fill the place of an element. In the theory of grammar, only units are accounted for structure while the elements can only be filled by an unit (see Figure 1.2). Allowing multiple units to fill an element requires accounting at least for the *order* if not also for the relation between the filler units. The structure as it is described in theories of grammar by Halliday (Halliday 2002) and Fawcett (Fawcett 2000) is defined by the unit and not the element. There is no direct reference in the theory to the unit ordering. Instead, the order relation is accounted in the structure through the concept of place. A unit has a specific possible structure in terms of places of elements which hold absolute position in the unit structure or relative one to each other. Therefore if an element is filled by two units simultaneously it constitutes a violation of the above principle as the order of those units is not accounted for but it matters which is easy to show in the following examples.

(18)   (Both my wife and her friend) arrived late.

(19)   * (And her friend both my wife) arrived late.

(20)   I want the front wall (either in blue or in green).

(21)   * I want the front wall (or in green either in blue).

If the order would not have mattered then we could say that the conjunctions from the example 18 can be reformulated into 19 and the one from 20 into 21. But such reformulations are grammatically incorrect. Obviously the places do matter and they need to be accounted in the unit structure as one element per place with no more than a single unit filling it.

I turn now to the role and position of lexical items signalling the conjunction which I consider having no place in the structure of the conjuncted units but outside of them,

that way forming together a higher order unit, the *complex unit*. This is contrary to what is being described in Cardiff and Sydney grammars for different reasons.

Fawcett present the Linker elements (&) which are filled by conjunctions as parts of virtually any unit class placed in the first position of the unit. For example in the "or in green" the presence of "or" signals the presence at least of one more unit of the same nature and does not contribute to the meaning of the prepositional group but to the meaning outside the group requiring presence of a sibling. Even more, the lack of a sibling most of the time would constitute an ungrammatical formulation. The only potential objection here is for the perfectly acceptable cases of clauses/sentences starting with a conjunction such as "and" or "but". In those cases the conjunction plays a textual function and still invites the presence of a sibling clause/sentence preceding the current one to be resolved in a clause complex or discourse level.

Halliday omits to discuss in IFG (Halliday & Matthiessen 2013) the place of Linkers. He implicitly proposes the same as Fawcett through his examples of paratactic relations at various rank levels (Halliday & Matthiessen 2013: 422, 534, 564, 566) that the lexical items signalling conjunction are included in the units they precede in the logical structure but not the experiential one. The main insufficiency here is that the logical structure does not provide any meaningful elements or unit class but some sort of proto-elements that resemble rather places than any functions. In this sense I consider treatment of conjunctions insufficiently accounted in IFG.

So conjunctions and pre-conjunctions shall not be placed inside the conjuncted units because they do not contribute to their meaning. They shall be enclosed as Linkers into unit complexes. But if we adopt the unit complexing then we need to define a unit structure. Hence I propose the following generic structure for the *coordination unit*.

Pre-Linker	Initiating Conjunct	... Conjunct ...	Linker	Conjunct
	1	$+ 2 ... + n_{-1}$		$+ n$

Table 1.11 Generic structure of the coordination unit

In Table 1.11 the first row presents a series of Conjuncts where the first one is initiating or the head and the rest are continuation Conjuncts of the former. In the first place there may be a Pre-Linker element such as "both" or "either" for example but it is optional and in the place before the last one is located the Linker element that determines the type of coordination. On the second row I provide, for orientation purposes, the Sydney logical structure of a paratactic expansion applied to the coordination unit complex. Note that the Pre-Linker and the Linker elements are merged with the conjuncted units.

Applying this structure to the previous example yields analysis such as in Table 1.12. The nominal group has the Epithet element filled by a coordination group formed of two Conjuncts and a Linker.

the	immediate		and	not	so	far	distant	future
Determiner	Epithet							Head
	Initiating Conjunct		Linker	Conjunct				

Table 1.12 Example analysis with coordination unit complex structure

Adopting the unit complex and in particular coordination unit requires two more clarifications (1) does the complex unit carry a syntactic class, and if so according to which criteria is it established? (2) Does it have any intrinsic features or all of them are inherited from the conjuncts?

Zhang states in her thesis that the coordinating constructions do not have any categorial features thus there is no need to provide a new unit type. Instead the categorial properties of the conjuncts are transferred upwards (Zhang 2010). For example if two nominal groups are conjuncted then the complex receives the nominal class.

This principle holds for most of the cases however there are rare cases when the units are of different classes. Consider 22 analysed in Table 1.13 where the conjuncts are a nominal group "last Monday" and a prepositional group "during the previous weekend".

(22)   I lost it (either last Monday or during the previous weekend).

either	last	Monday	or	during	the	previous	weekend
Pre-Linker	Initiating Conjunct		Linker	Conjunct			

Table 1.13 Coordination group with mixed class conjuncts

In this case there are two unit types that can be raised and it is not clear how to resolve this case. Options are (a) to leave the generic class *coordination complex*, (b) transfer the class of the first unit upwards, or (c) semantically resolve the class as both represent temporal circumstances even if they are realised through two different syntactic categories. This means that if no sub-classification is provided based on the constituent units below than there is no need to project/transfer upward the class of

the conjunct units. In this work I decided to leave the class generic and leave for the future an extensive unit complex classification.

I turn now to the last issue of this discussion, specifically whether the complex unit may have intrinsic features emerging from the conjunct elements.

In the example 23 the conjunction of two singular noun groups requires plural agreement with the verb. Even though semantic interpretation that only one item is selected at a time, syntactically both items are listed in the clause and attempting third person singular verb forms in 24 is grammatically incorrect. That leads to the conclusion that the coordination complex can have categorial features which none of the constituting units has.

(23)   A pencil or a pen **are** equally good as a smart-phone.

(24)   * A pencil or a pen **is** equally good as a smart-phone.
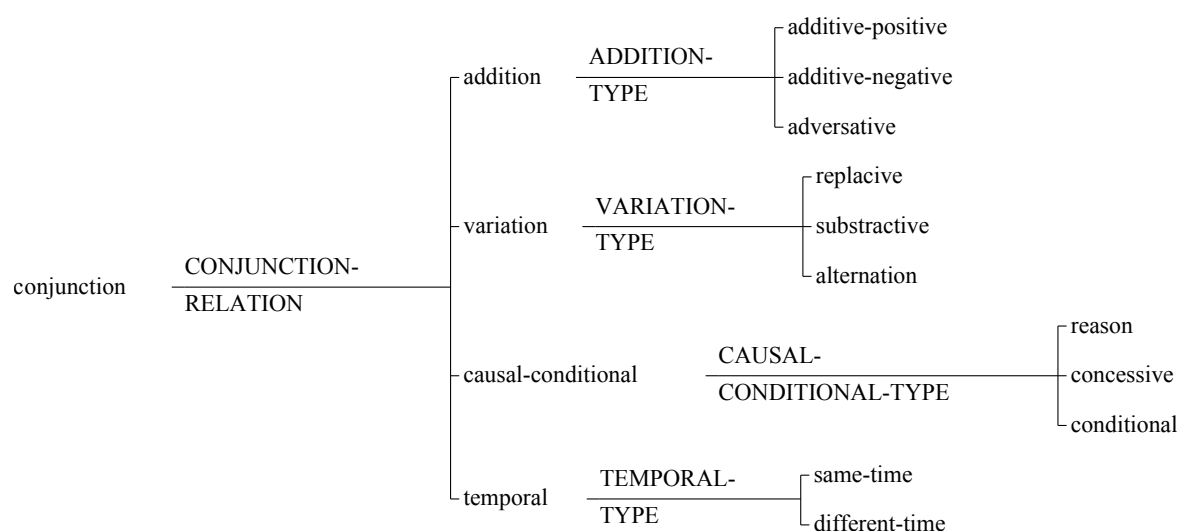
Fig. 1.11 Systemic network of coordination types

In the case of nominal group conjunction we can see that the plural feature emerges even if each individual unit is singular. For other unit classes it is not so obvious whether there are any linguistic features that emerge at the conjunction level. The meaning variation is semantic as for example conjunction of two verbs or clauses might mean very different things such as consecutive actions, concomitant actions or presence of two states at the same time and so on. This brings us to another feature of the coordination complex - the type of the relationship it constructs. The lexical items expounding the Linker and Pre-Linker (e.g. *and*, *or*, *but*, *yet*, *for*, *nor* or *so*) are indicator of relationship among the conjuncts and together can be systematised as the relationship types in the systemic network in Figure 1.11.

This section laid out how and why I treat the coordination phenomena in parsing. I adopt the unit complexing mechanism with taxis relations described in Sydney grammar in order to account for a new unit class, the *coordination unit*. I do that to ensure that each element of a unit is filled by no more than one other unit contrary to what Cardiff grammar proposes (see Definition 1.3.7). But taxis relations in Sydney grammars are represented via logical structures which is not rich enough to account for internal structure of the coordination unit. Therefore I also propose here a unit structure in terms of ordered functional elements just as for the rest of unit classes.

1.5 The grammatical units for parsing

Now that the important theoretical details have been covered I would like focus on the specific grammars of the two schools. They have common parts and also differ in large parts on their paradigmatic and syntagmatic descriptions. This section discusses the main units considered in each of the grammars. As in the previous section I argue on pragmatic grounds for the adoption of unit structures from one or the other grammar. The argument runs along the lines that some unit structures are closer to the traditional syntactic analysis and thus are easier to detect and parse and the other ones may be a level of abstraction higher falling more on semantic grounds and thus becoming more difficult to capture in structural variance and requiring lexico-semantic resources.

1.5.1 Verbal group and clause boundaries

In Sydney Grammar the verbal group is described as an expansion of a verb just like the nominal group is the expansion of the noun(Halliday & Matthiessen 2013: 396). There are certainly words that are closely related and syntactically dependent on the verb all together forming a unit that functions as a whole. For example the auxiliary verbs, adverbs or the negation particles are words that are directly linked to a lexical verb. The verb group functions as Finite + Predicator elements of the clause in Mood structure and as Process in Transitivity structure.

In Cardiff Grammar the verb group is dissolved moving the Main Verb as the pivotal element of the Clause unit. All the elements that form the clause structure and those that form the verb group structure are brought up together to the same level as elements of a clause. The clause structure in Cardiff Grammar comprises elements with clause related functions(like Subject, Adjunct, Complement etc.) and other elements with Main Verb related functions(Auxiliary, Negation particle, Finite operator etc.).

Regarded from the Hallidayan rank scale perspective, merging the elements of the verb group into clause structure is not permitted because the units are at different ranks. However it is not a problem for the relaxed rank scale version presented in subsection 1.4.1. The reason for adopting such an approach is best illustrated via complex verb groups with more than one non-auxiliary verb such as in examples 25–27.

I begin by addressing the impact of this merger on (a) the clause structure (b) the clause boundaries and (c) semantic role distribution within the clause.

(25)   (The commission **started to investigate** two cases of overfishing in Norway.)

(26)   (The commission **started** (**to investigate** two cases of overfishing in Norway.))

(27)   (The commission **started** (**to finish** (**investigating** two cases of overfishing in Norway.)))

In Sydney Grammar "started to investigate" (in 25) is considered a single predicate of investigation which has specified the aspect of event incipiency despite the fact that there are two lexical verbs within the same verbal group. The "starting" doesn't constitute any kind of process in semantic terms but rather specifies aspectual information about the investigation process. This is argued by looking at the conditions on the participants and it is equivalent in a formal approach to looking at where the selection restrictions for complements come from. The boundaries of the clause governed by this predicate stretch to the entire sentence.

Semantically it is a sound approach because despite the presence of two lexical verbs there is only one event. However allowing such compositions leads to unwanted syntactic analysis for multiple lexical verb cases in examples such as 27. To solve this kind of problem Fawcett dismisses the verb groups and merges their elements into clause structure. He proposes the syntactically elegant principle of *one main verb per clause* (Fawcett 2008). Applying this principle to the same sentence yields a structure of two clauses illustrated in example 26 where the main clause is governed by the verb "to start" and the embedded one by the verb "to investigate". Note the conflict between "one main verb per clause" with Halliday's principle that only whole units form the constituency of others (the (c) principle of rank scale described in subsection 1.4.1). So allowing incomplete groups into the constituency structure would breach the entire idea of unit based constituency.

Semantically the clause in SFL is a description of an event or situation as a figure with a process, participants and eventually circumstances where the process is realised through a lexical verb. Looking back to our examples, does the verb "to start" really describes a process or merely an aspect of it? Halliday treats such verbs

as aspectual and when co-occurring with other lexical verbs they are considered to form a single predicate. Accommodating Fawcett's stance, mentioned above and contradicting Halliday's approach, requires weakening the semantic requirement and allowing aspectual verbs to form clauses that contribute *aspectually or modally* to the embedded ones. I mention also the modal contribution because some verbs like *want, wish, hope* and others behave syntactically like the aspectual ones. Moreover, Fawcett introduces into the Cardiff Grammar Transitivity network an *influential* process type including all categories of meanings that semantically function as process modifiers: tentative, failing, starting, ending etc.

I adopt here Fawcett's "one main verb per clause" principle which as a consequence changes the way clauses are partitioned, leads to abolition of the verbal group and introduces the "influential" process type. Next I discuss it's impact on the structure of the clause unit.

1.5.2   The Clause

It is commonly agreed in linguistic communities that the unit of the clause is one of the core elements in human language. The main clause constituents are roughly the same in SFL as the ones in the traditional grammar (Quirk et al. 1985), transformational grammar (Chomsky 1957a) an indirectly in dependency grammar (Hudson 2010).

In current work I adopt the Cardiff Grammar clause structure with the *Main Verb* as pivotal element. Though there is no element that is obligatorily realised in English, I consider in the current work, the Main Verb to be way to flag a clause. In SFL are described clauses without a main verb such as minor clauses (exclamations, calls, greetings and alarms) that occur in conversational contexts and elliptical clauses Halliday & Matthiessen (2013) such as the one in example 28, none of which are covered in the present work.

(28)   They were in the bar, *Dave in the restroom and Sarah by the bar.*

I adopt Cardiff Grammar clause structure for English. It is formed of the *Subject, Finite, Main Verb*, up to two *Complements* and a various number of *Adjuncts*. All the elements of the assumed verbal group are part of the clause as well such as Auxiliary Verbs, Main Verb Extensions, Negators etc. (see Appendix **??** for a complete list).

1.5.3 The Nominal Group

The nominal group expresses things, classes of things or a selection of instances in that class. This section argues for adoption of the Sydney grammar noun group structure with a specific extension. It is that the elements of the nominal group can be filled, in addition to word classes, by the groups as well. This possibility is opened by the rank scale relaxation (Section 1.4.1) and Cardiff embedding principle (Definition 1.3.9). In addition to that I argue below for working on semantic and syntactic heads in two steps. First create the structure with the syntactic one (the Head) and then derive the semantic one (the Thing).

those	*two*	*old*	*electric*	*trains*	*from Luxembourg*
Pre-Modifier				Head	Post-Modifier
Deictic	Numerative	Epithet	Classifier	Thing	Qualifyier
determiner	numeral	adjective	adjective	noun	prepositional phrase

Table 1.14 An example of a nominal group in the Sydney Grammar (Halliday & Matthiessen 2013: 264)

In Table 1.14 an example analysis is presented of the nominal group proposed in the Sydney grammar (Halliday & Matthiessen 2013: 364–369). The Sydney nominal group is constituted by a head nominal item modified by descriptors or selectors such as: *Deictic*, *Numerative*, *Epithet*, *Classifier*, *Thing* and *Qualifier*. Each element has a fairly stable correspondence to the word classes, expected to be expounded by lexical items. Table 1.15 presents the mappings between the elements of nominal group and the word classes.

Experiential function in noun group	**class (of word or unit)**
Deictic	determiner, predeterminer, pronoun, adjective
Numerative	numeral(ordinal or cardinal)
Epithet	adjective
Classifier	adjective, noun
Thing	noun
Qualifier	prepositional phrase, clause

Table 1.15 Mapping of noun group elements to classes (Halliday & Matthiessen 2013: 379)

Inspired from Cardiff grammar, in addition to word classes, the elements of the nominal group can also be filled by the group classes corresponding to each word class

above. This way the Numerative, in addition to words, can be filled by a noun group, Epithet by an adjectival group, Classifier by an adjective or noun group and finally each of the elements can be filled by a coordination group discussed in Section 1.4.6.

The elements in Cardiff Grammar differ from those of Sydney Grammar. Table 1.16 exemplifies a noun group analysed with Cardiff Grammar covering all the possible elements. Table 1.17 provides a legend for the Cardiff Grammar acronyms along with mappings to unit and word classes that can fill each element.

or	*a photo*	*of*	*part*	*of*	*one*	*of*	*the best*	*of*	*the*	*fine*	*new*	*taxis*	*in Kew*	*,*
				pre-modifiers								head	post-modifiers	
&	rd	v	pd	v	qd	v	sd or od	v	dd	m	m	h	q	e

Table 1.16 The example of a nominal group in Cardiff Grammar

sym-bol	function meaning	class (of word or unit)
rd	representational determiner	noun, noun group
v	selector "of"	preposition
pd	partitive determiner	noun, noun group
fd	fractional determiner	noun, noun group, quantity group
qd	quantifying determiner	noun, noun group, quantity group
sd	superlative determiner	noun, noun group, quality group, quantity group
od	ordinative determiner	noun, noun group, quality group
td	tipic determiner	noun, noun group
dd	deictic determiner	determiner, pronoun, genitive cluster
m	modifier	adjective, noun, quality group, genitive cluster
h	head	noun, genitive cluster
q	qualifier	prepositional phrase, clause
&	linker	conjunction
e	ender	punctuation mark

Table 1.17 The mapping of noun group elements to classes in Cardiff grammar

The elements in Cardiff Grammar are based on semantic criteria supported by lexical and syntactic choices. Consequently some elements cannot be derived based on solely syntactic criteria, requiring semantically motivated lexical resources. Semantically bound elements which are a challenge are predominantly determiners *Representational, Partitive, Fractional, Superlative, Typic Determiners* while the rest of the elements:

Head, Qualifier, Selector, Modifier and Deictic, Ordinative and Quantifying Determiners can be determined solely on the syntactic criteria. The latter correspond fairly well to the Sydney version of nominal group which is adopted in the present work with the benefits of the relaxed rank system replacing the sub-structures with embedded units and simplifying the syntactic structures.

Another simplification is renouncing to distinction between the Head and Thing (Halliday & Matthiessen 2013: 390–396) discussed in Section 1.4.5. Thus if the logical Head of the nominal group is a noun then it is labelled as the Thing leaving the semantic discernment as a secondary process and out of the current scope. Otherwise, in cases of nominal groups without the Thing element, if the Head is a pronoun (other than personal), numeral or adjective (mainly superlatives) then they function as Deictic, Numerative or Epithet. So, as will be described in Chapter **??**, I propose to parse the nominal groups in two steps: first determine the main constituting chunks and assign functions to the unambiguous ones and second perform a semantically driven evaluation for the less certain units.

Next I explain the two step process using for illustration cases when the Thing is present but it is different from the Head such as in examples 29–31.

(29)   (a cup) of (tea)

(30)   (some) of (those youngsters)

(31)   (another one) of (those periodic eruptions)

These nominal groups can be analysed in two ways. Either being about the "cup", "some" or "another one" leading to a structure where the first noun is the head succeeded by a prepositional phrase Qualifier; or rather about "tea", "youngsters" and "eruptions" where the second noun is the head and so adopting a structure with complex determiners.

Table 1.18 shows on the first row an analysis with syntactic head i.e. the Head defined in Sydney grammar and on the second row an analysis with semantic head i.e. the Head defined in Cardiff grammar that also coincides with the Thing from Sydney grammar.

The syntactic Head is always the first noun in the nominal group. In the semantic evaluation phase special attention is given to Qualifiers filled by prepositional phrases starting with "of" preposition and whether the nominal group may function as qualifying, quantifying, ordination or other type of determiner.

Cardiff Grammar weakens the assumption that every prepositional phrase acts as Qualifier in a nominal group and it the special case of the preposition "of". It is

allowed to act not as the element introducing a prepositional phrase but as a end mark of a determiner-like selector. Thus making the former noun group a determiner in the latter one.

If the above conditions are satisfied, in the semantic evaluation phase, then the prepositional phrase Qualifyier is disassembled, the the preposition "of" is ascribed as a Selector element of the nominal group (in a way an upwards transfer) and the former nominal group (syntactically headed) becomes one of the determiners. This approach shifts the noun group head into the position of semantically based Thing and erases the discrepancy problem between them.

a	cup	of	tea
Determiner	Head	Qualifier	
Quantifying Determiner		Selector	Head/Thing

Table 1.18 Example analysis with syntactic and semantic heads

(32)   He was the **confidant** of the prime minister.

(33)   It was the **clash** of two cultures.

The above explanation is not a straight forward solution. The distinction between cases when the proposition "of" introduces a Qualifier or ends a Selector/Deictic requires lexical-semantic informed decision answering the question "what is the Thing that this nominal group is about?". And there is a lot of space for variations the syntactic structure. For example in 32 (Head/Thing marked in bold) the preposition "of" introduces Qualifiers.

In this section was discussed in detail the problem of semantic and syntactic heads (started in Section 1.4.5) applied to nominal groups in particular and how to approach parsing them. I conclude that it is fairly unambiguous and straight forward to determine the structure of nominal groups according to Sydney grammar yielding a syntactically founded structure. Once such a structure is ready it serves a a proper basis for a semantic evaluation that can be performed in terms of Cardiff grammar resulting in potential restructuring of the nominal group. The current implementation of the parser only the generation of syntactic structure is implemented leaving the semantically motivated noun groups for the future works.

1.5.4   The Adjectival and Adverbial Groups

This section introduces how the adverbial and adjectival groups are handled by the Sydney grammar and then how their equivalent quality group is structured in the

Cardiff grammar. As the structure of the quality group is semantically motivated some elements may be identified still at the syntactic level whereas some other ones require a more sophisticated lexical-semantic resources. In the last part of the section I estimate the complexity of parsing some the quality group elements elements.

Following the rationale of head-modifier similar to the case of nominal groups, the adjectives and adverbs function as pivotal elements to form groups. The structure of adverbial and adjectival constructions is briefly covered in the Sydney grammar in terms of head-modifier logical structures without an elaborated experiential structure like in the case of nominal groups. While the adverbial group is recognised as a distinct syntactic unit, the adjectival group is treated as a special case of nominal group.

(34)   You're *a very lucky boy.*

(35)   You're *very lucky.*

(36)   *The very lucky (one)* is you.

In the environments where nominal group functions as Attribute, typically in the attributive clauses such as 34, it can take also more contracted forms without the Thing and Deictic where the Head moves left onto the Epithet such as in example 35. One particularity of these nominal groups which here are distinguished as *adjectival group* units is that they cannot function as subject. For the example 36 to be grammatical, where the Attribute is in the Subject position, I had to add a determiner and eventually an unspecified nominal Head.

The *adverbial group*, in Sydney Grammar, has an adverb as Head which may or may not be accompanied by modifying elements (Halliday & Matthiessen 2013: 419). The adverbial groups may fill modal and circumstantial adjunct elements in a clause corresponding to eight semantic classes of: time, place, four types of manner and two types of assessment. The adverbial pre-modifiers express polarity, comparison and intensification along with only one comparison post-modifier (Halliday & Matthiessen 2013: 420–421). The adjectival and adverbial group are covered by the *quality group* unit in Cardiff grammar.

A thorough systemic functional examination in terms of lexis has been provided for the first time by Tucker Tucker (1997, 1998) materialised into a lexical-grammatical systematization of adjectives and the fine grained structure of quality group. He avoids calling the group according to the word class (adjective or adverb) but rather refers to the semantic meaning of what both groups express, i.e. the quality of things, situations or qualities themselves. The qualities of things have adjectives as their head while the qualities of situations an adverb.

In Cardiff Grammar, the head of the quality group is called *Apex* while the set of modifying elements: *Quality Group Deictic, Quality Group Quantifier, Emphasizing Temperer, Degree Temperer, Adjunctival Temperer, Scope* and *Finisher*. The quality group most frequently fills complements and adjuncts in clauses and fill modifiers and superlative determiners in nominal groups but there are also other cases found in the data.

Just like in the case of nominal group the adverbial and adjectival groups in Cardiff grammar are semantically motivated. To automatically identify elements of the quality group would according to this scheme therefore require lexico-semantic resources.

I turn now to discuss some relevant affinities concerning the adverbial groups. Some adverbs are different from others at least because not all of them can be heads of the adverbial group. Usually the adverbs that cannot act as heads, such as for example *very, much, less, pretty*, function as Emphasizing and Degree Temperers. The same ones also act as adjectival modifiers. A naive attempt to identify these Temperers would be to use a list of frequent words found in these functions.

Other elements of the quality group like the *Scoper* or *Finisher* are more difficult to identify and localize as part of the group only by syntactic cues and/or lists of words because of their inherent semantic nature. The problem is similar to detecting whether a prepositional phrase is filling a qualifier element in the preceding nominal group or is filling a complement or adjunct in the clause. Not surprisingly the Scopers and Finishers are most of the time prepositional phrases.

Another issue is continuity. The question is whether a grammar shall allow at least at a syntactic level discontinuous constituents or not. And then if so how to detect all the parts of the group even if they do not stand in proximity to each other. For example, comparatives, a complex case of a quality group, could be realised in a continuous or discontinuous forms. Compare the analyses presented in Table 1.19 and 1.20. In the first case the comparative structure is a continuous quality group. In the second case the comparative is dissociated and analysed as separate adjuncts.

On one hand it is not a problem treating them as two adjuncts, because that is what they are from the syntactic point of view. However, semantically as Fawcett proposes, there is only one quality group with a discontinuous realisation whose Scope element is placed in a thematic position before the Subject.

I	*am*	*much*	*smarter*	*today*	*than*	*yesterday*
Subject	Main Verb	Adjunct				
pronoun	verb	quality group				
		Emphasizing Temperer	Apex	Scope	Finisher	

Table 1.19 Comparative structure as one quality group adjunct

Today	*I*	*am*	*much*	*smarter*	*than*	*yesterday*
Adjunct	Subject	Main Verb	Adjunct			
adverb	pronoun	verb	quality group			
			Emphasizing Temperer	Apex	Finisher	

Table 1.20 Comparative structure split among two adjuncts

For an automatic process to identify a complex quality group is a difficult task. It needs to pick up cues like a comparative form of the adjective followed by the preposition "than" and then look for two terms being compared. Given some initial syntactic structure such patterns could be modelled and applied but only as a secondary semantically oriented process.

Since both the adverbial and adjectival groups have similar structures, it is syntactically feasible to automatically analyse them in terms of head-modifier structures in a first phase followed by a complementary process which assigns functional roles to the quality group components.

1.6   Selected system networks for parsing

The previous section described the units of structure adopted in the current work. I turn now to describe a few systemic networks relevant to selected units that are fairly low in delicacy and are quite close to what is considered grammatical features in the traditional grammar.

It is close to terms and concepts of syntactic structure in traditional grammar (role and definition of Subject, Complement and other functional elements) and the .

1.6.1   MOOD

One of the first system networks presented in Introduction to Functional Grammar (Halliday & Matthiessen 2013) is that of MOOD. This system is introduced in the

discussion of interpersonal metafunction of language. In this view a clause is conceptualised as message exchanged between dialogue interactants. A range of grammatical features from traditional grammar such as *mood*, *modality*, *aspect*, *mode*, *polarity*, *tense* etc. are covered, under different names, in this system network.

The terms in SFL literature sometimes are capitalised to distinguish system names, functions and features. Note that here the *MOOD* (all capital) refers here to the name of the system network; the *Mood* (first capital) is an element of clause structure formed of the Subject and Finite elements which, in fact, is not used in this work because of a general orientation towards Cardiff approach to structure; and the *mood* (no capital) is a type of feature carried by finite clauses (e.g. imperative, indicative, interrogative etc.)

Figure 1.12 presents

1.6.2   TRANSITIVITY

1.7   Discussion

This chapter has introduced the fundamentals of systemic functional linguistics and presented a consideration of Sydney and Cardiff theories of grammar to the task of parsing.

Because of its bottom up approach to unit structure, rank scale relaxation and accommodation of embedding as a general principle, Cardiff systemic functional theory is more suitable for parsing than the Sydney one. Nonetheless the unit definitions in the Cardiff grammar are deeply semantic in nature. Parsing with such units requires most of the time lexical-semantically informed decisions beyond merely syntactic variations. This is one of the reasons why the parsing attempts by O'Donoghue (1991) and others in the COMMUNAL project were all based on a corpus.

As there was no corpus available and because the parsing approach is based on a syntactic backbone none of the theories could be fully used as such. Section 1.4 and 1.5 attempts to merge and adapt the grammars and theories of grammar to the parsing approach of this thesis.

Next chapter lays the theoretical foundations of Dependency Grammar and introduces the Stanford dependency parser used as a departing point in current parsing pipeline. Because there is a transformation step from dependency to systemic functional consistency structure, the next chapter also covers a theoretical compatibility analysis and how such a transformation should in principle look like.
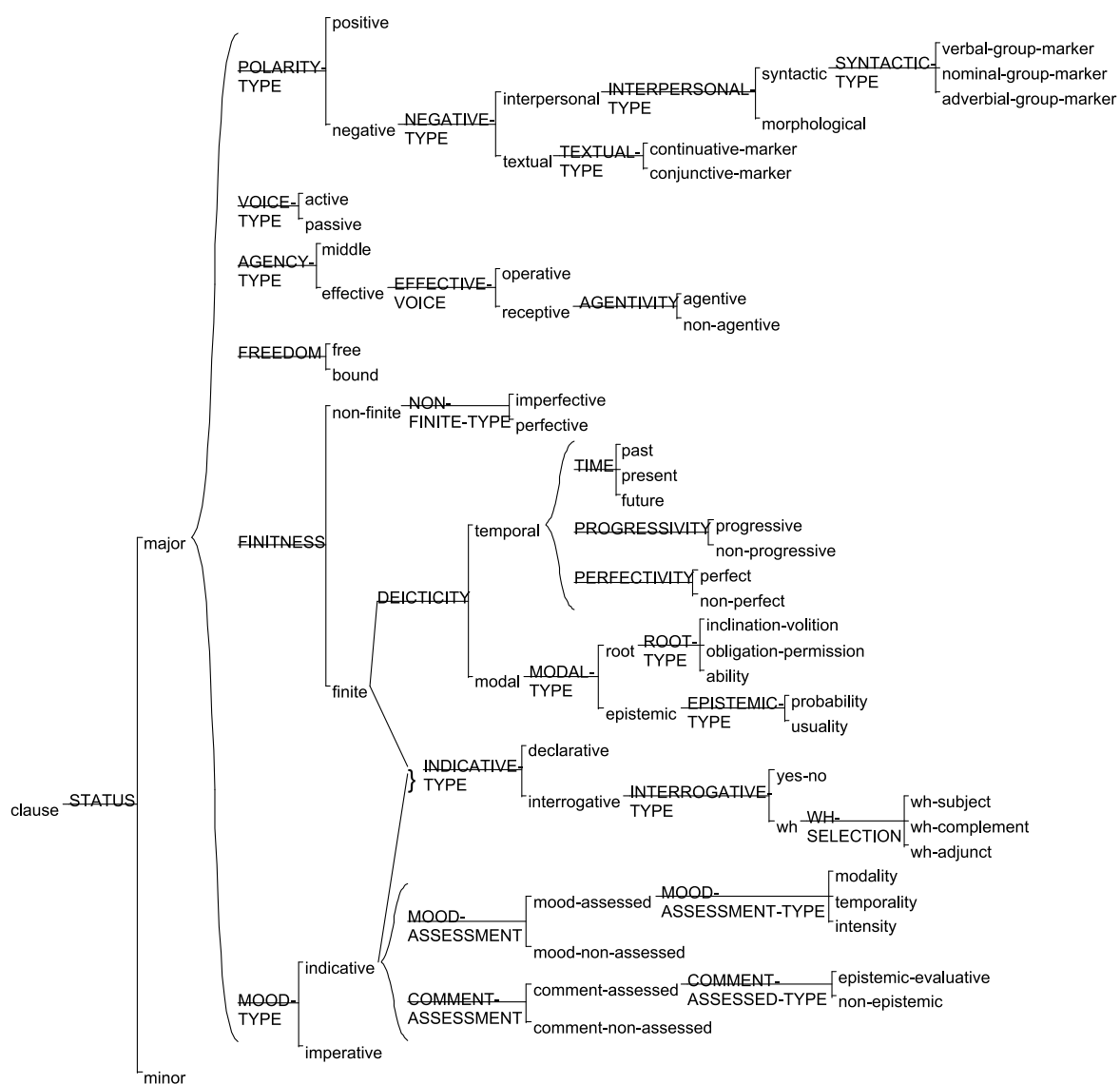
Fig. 1.12 An adaptation of the MOOD system network (Halliday & Matthiessen 2013: 162)

References

Abney, S. 1987. *The English noun phrase in its sentential aspects.* MIT Press.

Austin, J L. 1975. *How to do things with words*, vol. 3 (Syntax and Semantics 1). Harvard University Press.

Bach, Emmon. 1966. *An introduction to transformational grammars.* Holt, Rinehart and Winston. Inc.

Baker, Collin F, Charles J Fillmore & John B Lowe. 1998. The Berkeley FrameNet Project. In Christian Boitet & Pete Whitelock (eds.), *Proceedings of the 36th annual meeting on association for computational linguistics*, vol. 1 ACL '98, 86–90. University of Montreal Association for Computational Linguistics. doi:10.3115/980845.980860. <http://portal.acm.org/citation.cfm?doid=980845.980860>.

Bar-Hillel, Yehoshua. 1953. A quasi-arithmetical notation for syntactic description. *Language* 29. 47–58. Reprinted in Y. Bar-Hillel. (1964). *Language and Information: Selected Essays on their Theory and Application*, Addison-Wesley 1964, 61–74.

Bateman, John A. 2008. Systemic-Functional Linguistics and the Notion of Linguistic Structure: Unanswered Questions, New Possibilities. In Jonathan J. Webster (ed.), *Meaning in context: Implementing intelligent applications of language studies*, 24–58. Continuum.

Bateman, John A, Renate Henschel & Fabio Rinaldi. 1995. The Generalized Upper Model . Tech. rep. GMD/IPSI. <http://www.fb10.uni-bremen.de/anglistik/langpro/webspace/jb/gum/gum-2.pdf>.

Bateman, John A, Joana Hois, Robert Ross & Thora Tenbrink. 2010. A linguistic ontology of space for natural language processing. *Artificial Intelligence* 174(14). 1027–1071. doi:10.1016/j.artint.2010.05.008. <http://linkinghub.elsevier.com/retrieve/pii/S0004370210000858>.

Bateman, John A. & Christian M. I. M. Matthiessen. 1988. Using a functional grammar as a tool for developing planning algorithms — an illustration drawn from nominal group planning. Tech. rep. Information Sciences Institute Marina del Rey, California. (Penman Development Note).

Bloomfield, Leonard. 1933. *Language.* New York and London: Henry Holt and Co. and Allen and Unwin Ltd.

Böhmová, Alena, Jan Hajič, Eva Hajičová & Barbora Hladká. 2003. *The prague dependency treebank* 103–127. Dordrecht: Springer Netherlands. doi:10.1007/978-94-010-0201-1_7. <https://doi.org/10.1007/978-94-010-0201-1_7>.

Bohnet, Bernd. 2010. Very high accuracy and fast dependency parsing is not a contradiction. In *Proceedings of the 23rd international conference on computational linguistics* COLING '10, 89–97. Stroudsburg, PA, USA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=1873781.1873792>.

Bresnan, Joan. 1982. Control and complementation. *Linguistic Inquiry* 13(3). 343–434.

Bresnan, Joan. 2001. *Lexical-Functional Syntax.* Blackwell. <http://books.google.lu/books/about/Lexical{_}Functional{_}Syntax.html?id=vMxgevXoq{_}gC{&}redir{_}esc=y>.

Buchholz, Sabine & Erwin Marsi. 2006. Conll-x shared task on multilingual dependency parsing. In *Proceedings of the tenth conference on computational natural language learning* CoNLL-X '06, 149–164. Stroudsburg, PA, USA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=1596276.1596305>.

Bühler, Karl. 1934. *Sprachtheorie: die Darstellungsfunktion der Sprache.* Jena: Fischer.

Butler, Christopher. 1985. *Systemic linguistics: Theory and applications.* Batsford Academic and Educational.

Butler, Christopher S. 2003a. *Structure and function: A guide to three major structural-functional theories; Part 1: Approaches to the simplex clause.* Amsterdam and Philadelphia: John Benjamins.

Butler, Christopher S. 2003b. *Structure and function: A guide to three major structural-functional theories; Part 2: From clause to discourse and beyond.* Amsterdam and Philadelphia: John Benjamins.

Carl Pollard & Ivan Sag. 1987. *Information-Based Syntax and Semantics.* CSLI.

Carreras, Xavier. 2007. Experiments with a higher-order projective dependency parser. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (emnlp-conll)*, .

Carreras, Xavier & Lluís Màrquez. 2005. Introduction to the conll-2005 shared task: Semantic role labeling. In *Proceedings of the ninth conference on computational natural language learning* CONLL '05, 152–164. Stroudsburg, PA, USA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=1706543.1706571>.

Carroll, John, Ted Briscoe & Antonio Sanfilippo. 1998. Parser evaluation: a survey and a new proposal. In *Proceedings of the international conference on language resources and evaluation*, 447–454.

Carroll, John, Guido Minnen & Ted Briscoe. 1999. Corpus Annotation for Parser Evaluation. In *Proceedings of the eacl workshop on linguistically interpreted corpora (linc)* June, 7. Association for Computational Linguistics. <http://arxiv.org/abs/cs/9907013>.

Cer, Daniel D.M., Marie-Catherine M.C. De Marneffe, Daniel Jurafsky & C.D. Manning. 2010. Parsing to stanford dependencies: Trade-offs between speed and accuracy. In *Lrec 2010*, vol. 0, European Language Resources Association. <http://171.64.67.140/pubs/lrecstanforddeps{_}final{_}final.pdfhttp://www.lrec-conf.org/proceedings/lrec2010/pdf/730{_}Paper.pdf>.

Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory* 2. 113–124.

Chomsky, Noam. 1957a. *Syntactic Structures.* Mouton & Co.

Chomsky, Noam. 1957b. *Syntactic structures.* The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the theory of syntax.* Cambridge, Massachusetts: M.I.T. Press.

Chomsky, Noam. 1981. *Lectures on Government and Binding.* Dordrecht: Foris.

Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding*, vol. 6. MIT press.

Chomsky, Noam. 1986. *Barriers*, vol. 13. MIT press.

Chomsky, Noam. 1993a. A Minimalist Program for Linguistic Theory. In K. Hale & S. J. Keyser (eds.), *The View from Building 20*, Cambridge, Mass: MIT Press.

Chomsky, Noam. 1993b. *Lectures on government and binding: The pisa lectures* 9. Walter de Gruyter.

Clegg, Andrew B & Adrian J Shepherd. 2007. Benchmarking natural-language parsers for biological applications using dependency graphs. *BMC bioinformatics* 8(1). 24.

Cordella, L P, P Foggia, C Sansone & M Vento. 2001. An improved algorithm for matching large graphs. *3rd IAPRTC15 workshop on graphbased representations in pattern recognition* 219(2). 149–159. doi:10.1.1.101.5342. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.5342{&}rep=rep1{&}type=pdf>.

Cordella, L P, P Foggia, C Sansone & M Vento. 2004. A (sub)graph isomorphism algorithm for matching large graphs. doi:10.1109/TPAMI.2004.75. <http://www.ncbi.nlm.nih.gov/pubmed/15641723>.

Day, Michael David. 2007. *A Corpus-Consulting Probabilistic Approach to Parsing : the CCPX Parser and its Complementary Components*: Cardiff University dissertation.

Domingos, Pedro, Stanley Kok, Daniel Lowd & Hoifung Poon. 2010. Markov Logic. *Journal of computational biology a journal of computational molecular cell biology* 17(11). 1491–508. doi:10.1089/cmb.2010.0044. <http://www.ncbi.nlm.nih.gov/pubmed/21685052>.

Fawcett, R. 1988a. Language Generation as Choice in Social Interaction. In Zock, M. & G. Sabah (eds.), *Advances in natural language generation*, vol. 2, 27–49. Pinter Publishers. (Paper presented at the First European Workshop of Natural Language Generation, Royaumont, 1987).

Fawcett, Robin. 2000. *A Theory of Syntax for Systemic Functional Linguistics.* John Benjamins Publishing Company paperback edn.

Fawcett, Robin P. 1988b. What makes a 'good' system network good? In James D. Benson & William S. Greaves (eds.), *Systemic functional approaches to discourse*, 1–28. Norwood, NJ: Ablex.

Fawcett, Robin P. 1990. The COMMUNAL project: two years old and going well. *Network: news, views and reviews in systemic lingustics and related areas* 13/14. 35–39.

Fawcett, Robin P. 1993. The architecture of the COMMUNAL project in NLG (and NLU). In *The Fourth European Workshop on Natural Language Generation*, Pisa.

Fawcett, Robin P. 2008. *Invitation to Systemic Functional Linguistics through the Cardiff Grammar.* Equinox Publishing Ltd.

Fawcett, Robin P. 2009. How to Analyze Process and Participant Roles. In *The functional semantics handbook: Analyzing english at the level of meaning*, Continuum.

Fawcett, Robin P. 2011. A semantic system network for MOOD in English (and some complementary system networks).

Fellbaum, Christiane & George Miller (eds.). 1998. *WordNet: An electronic lexical database.* The MIT Press.

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6(2). 222–254. <http://scholar.google.it/scholar?q=fillmore{&}hl=it{&}btnG=Cerca{&}lr={#}5>.

Fillmore, Charles J, Christopher R Johnson & Miriam RL Petruck. 2003. Background to framenet. *International journal of lexicography* 16(3). 235–250.

Firth, J.R. 1957. A synopsis of linguistic theory 1930-1955. *Studies in linguistic analysis* 1–32. <http://www.bibsonomy.org/bibtex/25b0a766713221356e0a5b4cc2023b86a/glanebridge>.

Gaifman, Haim. 1965. Dependency systems and phrase-structure systems. *Information and Control* 8(3). 304 – 337. doi:https://doi.org/10.1016/S0019-9958(65)90232-9. <http://www.sciencedirect.com/science/article/pii/S0019995865902329>.

Gale, D. & L. S. Shapley. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly* 69(1). 9–15. <http://www.jstor.org/stable/2312726>.

Garde, Paul. 1977. Ordre lineaire et dependance syntaxique. contribution a une typologie. In *Bulletin de la societe de linguistique de paris*, vol. 1 72, 1–26. Paris.

Gerdes, Kim & Sylvain Kahane. 2013. Defining dependencies (and constituents). *Frontiers in Artificial Intelligence and Applications* 258. 1–25.

Gildea, Daniel & Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational linguistics* 28(3). 245–288.

Haegeman, Liliane. 1991. *Introduction to Government and Binding Theory*, vol. 2. Blackwell.

Hajic, Jan, Eva Hajicová, Petr Pajas, Jarmila Panevová, Petr Sgall & Barbora Vidová-Hladká. 2001. Prague dependency treebank 1.0 (final production label). cdrom cat: Ldc2001t10. Tech. rep. ISBN 1-58563-212-0.

Halliday, Michael A. K. 1957. Some aspects of systematic description and comparison in grammatical analysis. In *Studies in Linguistic Analysis*, 54–67. Oxford: Blackwell.

Halliday, Michael A. K. 1961a. Categories of the theory of grammar. *Word* 17(3). 241–292.

Halliday, Michael A. K. 1961b. Categories of the theory of grammar. *Word* 17(3). 241–292. Reprinted in abbreviated form in Halliday (1976) edited by Gunther Kress, pp 52-72.

Halliday, Michael A. K. 1994. *An Introduction to Functional Grammar*. London: Edward Arnold 2nd edn.

Halliday, Michael A. K. 1996. On grammar and grammatics. In Ruqaiya Hasan, Carmel Cloran & David Butt (eds.), *Functional descriptions – theory in practice* Current Issues in Linguistic Theory, 1–38. Amsterdam: Benjamins.

Halliday, Michael A. K. 1997. Linguistics as metaphor, 3–27. Continuum.

Halliday, Michael A. K. 2003a. Ideas about language. In Michael A. K. Halliday & Jonathan J. Webster (eds.), *On language and linguistics. Volume 3 of collected works of M.A. K. Halliday*, 490. New York: Continuum.

Halliday, Michael A. K. 2003b. Systemic theory. In Michael A. K. Halliday & Jonathan J. Webster (eds.), *On language and linguistics. Volume 3 of collected works of M.A. K. Halliday*, 490. New York: Continuum.

Halliday, Michael A.K. 2002. Categories of the theory of grammar. In Jonathan Webster (ed.), *On grammar (volume 1)*, 442. Continuum.

Halliday, Michael A.K. 2003c. On the "architecture" of human language. In Jonathan Webster (ed.), *On language and linguistics*, vol. 3 Collected Works of M. A. K. Halliday, 1–32. Continuum.

Halliday, Michael A.K. & Christian M.I.M. Matthiessen. 2013. *An Introduction to Functional Grammar (4th Edition)*. Routledge 4th edn.

Halliday, Michael A.K. & M.I.M. Matthiessen, Christian. 2004. *An introduction to functional grammar (3rd Edition)*. Hodder Education.

Harris, Z.S. 1951. *Methods in structural linguistics* Methods in Structural Linguistics. University of Chicago Press. <https://books.google.lu/books?id=a6nYjgEACAAJ>.

Harrison, Philip, Steven Abney, Ezra Black, Dan Flickinger, Claudia Gdaniec, Ralph Grishman, Donald Hindle, Robert Ingria, Mitch Marcus, Beatrice Santorini & Tomek Strzalkowski. 1991. Evaluating syntax performance of parser/grammars. In *Proceedings of the natural language processing systems evaluation workshop, berekely, ca, june 1991* Rome Laboratory Technical Report, RL-TR-91-362, .

Hasan, Ruqaiya. 2014. The grammarian's dream: lexis as most delicate grammar. In Jonathan Webster (ed.), *Describing language form and function*, vol. 5 Collected Works of Ruqaiya Hasan, chap. 6. Equinox Publishing Ltd.

Hays, David G. 1960. Grouping and dependency theories. *Proceedings of the National Symposium on Machine Translation* 2538. 258–266.

Hays, David G. 1964. Dependency theory: A formalism and some observations. *Language* 40(4). 511–525. <http://www.jstor.org/stable/411934>.

Hjelmslev, Louis. 1953. *Prolegomena to a theory of language*. Bloomington, Indiana: Indiana University Publications in Anthropology and Linguistics. Translated by Francis J. Whitfield.

Hockett, Charles F. 1958. *A course in modern linguistics*. New York: Macmillan.

Honnibal, Matthew. 2004. Converting the Penn Treebank to Systemic Functional Grammar. *Technology* 147–154.

Honnibal, Matthew & Jr James R Curran. 2007. Creating a systemic functional grammar corpus from the Penn treebank. *Proceedings of the Workshop on Deep . . .* 89–96. doi:10.3115/1608912.1608927. <http://dl.acm.org/citation.cfm?id=1608927>.

Hudson, Richard. 2010. *An Introduction to Word Grammar*. Cambridge University Press.

Hutchins, W John. 1999. Retrospect and prospect in computer-based translation. In *Proceedings of mt summit vii "mt in the great translation era"* September, 30–44. AAMT.

Johnson, Christopher & Charles J. Fillmore. 2000. The framenet tagset for frame-semantic and syntactic coding of predicate-argument structure. In *Proceedings of the 1st north american chapter of the association for computational linguistics conference* NAACL 2000, 56–62. Stroudsburg, PA, USA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=974305.974313>.

Kasper, Robert. 1988. An Experimental Parser for Systemic Grammars. In *Proceedings of the 12th International Conference on Computational Linguistics*, .

Kay, Martin. 1985. Parsing In Functional Unification Grammar. In D.Dowty, L. Karttunen & A. Zwicky (eds.), *Natural language parsing*, Cambridge University Press.

King, Tracy Holloway & Richard Crouch. 2003. The PARC 700 Dependency Bank. In *4th international workshop on linguistically interpreted corpora (linc03)*, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.15.3277>.

Kipper, Karin, Anna Korhonen, Neville Ryant & Martha Palmer. 2008. A large-scale classification of English verbs. *Language Resources And Evaluation* 42(1). 21–40. doi:10.1007/s10579-007-9048-2.

Koeva, Svetla, Borislav Rizov, Ekaterina Dimitrova, Tarpomanova Tsvetana, Rositsa Dekova, Ivelina Stoyanova, Svetlozara Leseva, Hristina Kukova & Angel Genov. 2012. Bulgarian-english sentence-and clause-aligned corpus. In *Proceedings of the second workshop on annotation of corpora for research in the humanities (acrh-2)*, 51–62.

Kübler, S., R. McDonald & J. Nivre. 2009. *Dependency parsing* Online access: IEEE (Institute of Electrical and Electronics Engineers) IEEE Morgan & Claypool Synthesis eBooks Library. Morgan & Claypool. <https://books.google.lu/books?id=k3iiup7HB9UC>.

Kucera, Henry & W. Nelson Francis. 1968. Computational Analysis of Present-Day American English. *American Documentation* 19(4). 419. doi:10.2307/302397. <http://search.ebscohost.com/login.aspx?direct=true{&}db=bth{&}AN=16865479{&}login.asp{&}site=ehost-live>.

Lecerf, Yves. 1961. Une representation algebrique de la structure des phrases dans diverses langues naturelles. In *Compte rendu de l'academie des sciences de paris* 252, 232–234. Academie des Sciences.

Lee, Jinsoo, Wook-Shin Han, Romans Kasperovics & Jeong-Hoon Lee. 2013. An in-depth comparison of subgraph isomorphism algorithms in graph databases. *Proceedings of the 39th international conference on Very Large Data Bases* 133–144. doi:10.14778/2535568.2448946. <http://dl.acm.org/citation.cfm?id=2448936.2448946>.

Lemke, Jay L. 1993. Discourse, dynamics, and social change. *Cultural Dynamics* 6(1-2). 243–276.

Lin, Dekang & Patrick Pantel. 2001. Discovery of inference rules for question-answering. *Natural Language Engineering* 7(4). 343–360.

Mann, William C. 1983a. An Overview of the PENMAN Text Generation System. Tech. Rep. ISI/RR-83-114 USC/Information Sciences Institute Marina del Rey, CA.

Mann, William C. 1983b. An overview of the PENMAN text generation system. In *Proceedings of the National Conference on Artificial Intelligence*, 261–265. AAAI. Also appears as USC/Information Sciences Institute, RR-83-114.

Mann, William C. & Christian M. I. M. Matthiessen. February 1983. A demonstration of the Nigel text generation computer program. In *Nigel: A Systemic Grammar for Text Generation*, USC/Information Sciences Institute, RR-83-105. This paper also appears in a volume of the *Advances in Discourse Processes Series*, R. Freedle (ed.): *Systemic Perspectives on Discourse: Volume I.* published by Ablex.

Mann, William C., Christian M. I. M. Matthiessen & Sandra A. Thompson. 1992. Rhetorical Structure Theory and Text Analysis. In William C Mann & Sandra A Thompson (eds.), *Discourse description: Diverse linguistic analyses of a fund-raising text*, vol. 16 Pragmatics & Beyond New Series, 39–79. John Benjamins Publishing Company.

Mann, William C & Sandra A Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text* 8(3). 243–281. doi:10.1515/text.1.1988. 8.3.243.

Marcus, Mitchell P, Beatrice Santorini & Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19(2). 313–330. doi:10.1162/coli.2010.36.1.36100. <http://portal.acm.org/citation. cfm?id=972470.972475>.

Marneffe, Marie-Catherine, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre & Christopher D. Manning. 2014. Universal Stanford dependencies: A cross-linguistic typology. In *Proceedings of the ninth international conference on language resources and evaluation (lrec-2014)(vol. 14)*, European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2014/pdf/ 1062{_}Paper.pdf>.

Marneffe, Marie-Catherine, Bill MacCartney & Christopher D Manning. 2006. Generating Typed Dependency Parses from Phrase Structure Parses. In *Lrec 2006*, vol. 6 3, 449–454. Stanford University. <http://nlp.stanford.edu/manning/papers/ LREC{_}2.pdf>.

Marneffe, Marie-Catherine & Christopher D. Manning. 2008a. Stanford typed dependencies manual. Tech. Rep. September Stanford University. <http://nlp.stanford. edu/downloads/dependencies{_}manual.pdf>.

Marneffe, Marie-Catherine & Christopher D. Manning. 2008b. The Stanford typed dependencies representation. *Coling 2008 Proceedings of the workshop on CrossFramework and CrossDomain Parser Evaluation CrossParser 08* 1(ii). 1–8. doi:10.3115/ 1608858.1608859. <http://portal.acm.org/citation.cfm?doid=1608858.1608859>.

Matthiessen, Christian M. I. M. 1995. *Lexicogrammatical cartography: English systems.* Tokyo, Taipei and Dallas: International Language Science Publishers.

Matthiessen, Christian M. I. M. & John A. Bateman. 1991. *Text generation and systemic-functional linguistics: experiences from English and Japanese.* London and New York: Frances Pinter Publishers and St. Martin's Press.

Matthiessen, M.I.M., Christian. 1985. The systemic framework in text generation: Nigel. In James Benson & Willian Greaves (eds.), *Systemic perspective on Discourse, Vol I*, 96–118. Ablex.

McCarthy, John, Marvin L. Minsky, Nathaniel Rochester & Claude E. Shannon. 2006. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine* 27(4). 12. doi:10.1609/aimag.v27i4.1904. <http://www.aaai.org/ojs/index.php/ aimagazine/article/view/1904{%}5Cnhttp://www.mendeley.com/catalog/ proposal-dartmouth-summer-research-project-artificial-intelligence-august-31-1955/ {%}5Cnhttp://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth. htmlhttp://>.

McDonald, David D. 1980. *Natural Language Production as a Process of Decision Making under Constraint*: MIT, Cambridge, Mass dissertation.

McDonald, Ryan, Koby Crammer & Fernando Pereira. 2005. Online large-margin training of dependency parsers. In *Proceedings of the 43rd annual meeting on association for computational linguistics*, 91–98. Association for Computational Linguistics.

McDonald, Ryan, Kevin Lerman & Fernando Pereira. 2006. Multilingual dependency analysis with a two-stage discriminative parser. In *Proceedings of the tenth conference on computational natural language learning* CoNLL-X '06, 216–220. Stroudsburg, PA, USA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=1596276.1596317>.

McDonald, Ryan & Fernando Pereira. 2006. Online learning of approximate dependency parsing algorithms. In *11th conference of the european chapter of the association for computational linguistics*, .

Mel'čuk, Igor. 1988. Semantic description of lexical units in an Explanatory Combinatorial Dictionary: basic principles and heuristic criteria. *International Journal of Lexicography* 1. 165–188.

Mel'čuk, Igor A. & Nikolaj V. Pertsov. 1986. *Surface syntax of english.* John Benjamins Publishing. doi:10.1075/llsee.13. <http://www.jbe-platform.com/content/books/9789027279378>.

Miller, George A. 1995. WordNet: a lexical database for English.

Miyao, Yusuke & Jun'ichi Tsujii. 2005. Probabilistic disambiguation models for wide-coverage hpsg parsing. In *Proceedings of the 43rd annual meeting on association for computational linguistics* ACL '05, 83–90. Stroudsburg, PA, USA: Association for Computational Linguistics. doi:10.3115/1219840.1219851. <https://doi.org/10.3115/1219840.1219851>.

Moravcsik, Edith A. 2006. *An Introduction to Syntactic Theory.* Continuum paperback edn.

Neale, Amy C. 2002. More Delicate TRANSITIVITY: Extending the PROCESS TYPE for English to include full semantic classifications. Tech. rep. Cardiff University.

Nivre, Joakim. 2006. *Inductive dependency parsing (text, speech and language technology).* Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Nivre, Joakim. 2015. Towards a universal grammar for natural language processing. In *Computational Linguistics and Intelligent Text Processing*, 3–16. Springer.

Nivre, Joakim, Johan Hall, Sandra Kübler, Ryan McDonald, Jens Nilsson, Sebastian Riedel & Deniz Yuret. 2007a. The conll 2007 shared task on dependency parsing. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (emnlp-conll)*, 915–932.

Nivre, Joakim, Johan Hall, Jens Nilsson, Atanas Chanev, Gülsen Eryigit, Sandra Kübler, Svetoslav Marinov & Erwin Marsi. 2007b. Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering* 13(2). 95–135. doi:10.1017/S1351324906004505. <https://doi.org/10.1017/S1351324906004505>.

Nivre, Joakim, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty & Daniel Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the tenth international conference on language resources and evaluation (lrec 2016)*, Paris, France: European Language Resources Association (ELRA).

O'Donnell, Michael. 1993. Reducing Complexity in Systemic Parser. In *Proceeedings of the third international workshop on parsing technologies*, .

O'Donnell, Michael. 1994. Sentence Analysis and Generation: a systemic perspective. Tech. rep. Department of Linguistics, University of Sydney.

O'Donnell, Michael. 2005. The UAM Systemic Parser. *Proceedings of the 1st Computational Systemic Functional Grammar Conference* <http://www.wagsoft.com/Papers/ODonnellUamParser.pdf>.

O'Donnell, Michael J. & John A. Bateman. 2005. SFL in computational contexts: a contemporary history. In Ruqiaya Hasan, M.I.M. Matthiessen, Christian & Jonathan Webster (eds.), *Continuing discourse on language: A functional perspective*, vol. 1 Booth 1956, 343–382. Equinox Publishing Ltd.

O'Donnell, Mick. 2008a. Demonstration of the UAM CorpusTool for text and image annotation. In *Proceedings of the acl-08:hlt demo session* June, 13–16.

O'Donnell, Mick. 2008b. The UAM CorpusTool: Software for Corpus Annotation and Exploration. In Bretones Callejas & Carmen M. (eds.), *Applied linguistics now: Understanding language and mind*, vol. 00, 1433–1447. Universidad de Almería.

O'Donoghue, Tim. 1991. The Vertical Strip Parser: A lazy approach to parsing. Tech. rep. School of Computer Studies, University of Leeds.

Pei, Wenzhe, Tao Ge & Baobao Chang. 2015. An effective neural network model for graph-based dependency parsing. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)*, vol. 1, 313–322.

Penman Project. 1989. PENMAN documentation: the Primer, the User Guide, the Reference Manual, and the Nigel Manual. Tech. rep. USC/Information Sciences Institute Marina del Rey, California.

Pollard, Carl J. & Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. University of Chicago Press.

Pollock, Jean-Yves. 1989. Verb movement, universal grammar, and the structure of ip. *Linguistic inquiry* 20(3). 365–424.

Postal, P. M. 1974. *On Raising*. MIT Press.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik & David Crystal. 1985. *A comprehensive grammar of the English language*, vol. 1 2. Longman. <http://www.amazon.com/dp/0582517346http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=2545152>.

Radford, Andrew. 1997. *Syntax: A Minimalist Introduction*. Cambridge University Press.

Richardson, Matthew & P. Domingos. 2006. Markov logic networks. *Machine learning* 62(1-2). 107–136. doi:10.1007/s10994-006-5833-1.

Saitta, Lorenza & Jean-Daniel Zucker. 2013. *Abstraction in artificial intelligence and complex systems*. Springer-Verlag New York. doi:10.1007/978-1-4614-7052-6. <http://www.springer.com/la/book/9781461470519>.

Santorini, Beatrice. 1990. Part-of-Speech Tagging Guidelines for the Penn Treebank Project (3rd Revision). *University of Pennsylvania 3rd Revision 2nd Printing* 53(MS-CIS-90-47). 33. doi:10.1017/CBO9781107415324.004. <http://www.personal.psu.edu/faculty/x/x/xxl13/teaching/sp07/apling597e/resources/Tagset.pdf>.

Saussure, Ferdinand de. 1959 [1915]. *Course in General Linguistics*. New York / Toronto / London: McGraw-Hill and the Philosophical Library, Inc. Edited by Charles Bally and Albert Sechehaye, in collaboration with Albert Riedlinger; translated by Wade Baskin.

Schank, Roger. 1969. Conceptual dependency as a framework for linguistic analysis. *Linguistics* 7(49). 28–50.

Schuler, Karin Kipper. 2005. Verbnet: A broad-coverage, comprehensive verb lexicon .

Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*, vol. 0. Cambridge University Press. <http://books.google.com/books?id=t3{_}WhfknvF0C{&}pgis=1>.

Shang, Haichuan, Ying Zhang, Xuemin Lin & Jeffrey Xu Yu. 2008. Taming verification hardness: an efficient algorithm for testing subgraph isomorphism. *VLDB Endow* 1(1). 364–375.

Silveira, Natalia, Timothy Dozat, Marie-Catherine De Marneffe, Samuel Bowman, Miriam Connor, John Bauer & Chris Manning. 2014. A Gold Standard Dependency Corpus for English. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the ninth international conference on language resources and evaluation (lrec'14)*, European Language Resources Association (ELRA). <http://nlp.stanford.edu/pubs/USD{_}LREC14{_}paper{_}camera{_}ready.pdf>.

Sleator, Daniel Dominic & David Temperley. 1995. Parsing english with a link grammar. *CoRR* abs/cmp-lg/9508004. 93. <http://arxiv.org/abs/cmp-lg/9508004>.

Snow, Rion, Daniel Jurafsky & Andrew Y Ng. 2005. Learning syntactic patterns for automatic hypernym discovery. In *Advances in neural information processing systems*, 1297–1304.

Souter, David Clive. 1996. *A Corpus-Trained Parser for Systemic-Functional Syntax*: University of Leeds Phd. <http://etheses.whiterose.ac.uk/1268/>.

Sowa, John F. 1976. Conceptual graphs for a data base interface. *IBM Journal of Research and Development* 20(4). 336–357. doi:10.1147/rd.204.0336. <http://dx.doi.org/10.1147/rd.204.0336>.

Stevan Harnad. 1992. The Turing Test Is Not A Trick: Turing Indistinguishability Is A Scientific Criterion. *SIGART Bulletin* 3(4). 9–10. <http://users.ecs.soton.ac.uk/harnad/Papers/Harnad/harnad92.turing.html>.

Stockwell, Robert P., Barbara Hall Partee & Paul Schacter. 1973. *The major syntactic structures of english.* New York: Holt, Rinehart and Winston. <http://hdl.handle.net/2027/mdp.39015002216417>. Bibliography: p. 811-840.

Stowell, T.A. & E. Wehrli. 1992. *Syntax and the lexicon* Syntax and semantics. Academic Press. <https://books.google.lu/books?id=yiEcAQAAIAAJ>.

Taverniers, Miriam. 2011. The syntax-semantics interface in systemic functional grammar: Halliday's interpretation of the Hjelmslevian model of stratification. *Journal of Pragmatics* 43(4). 1100–1126. doi:10.1016/j.pragma.2010.09.003.

Tesniere, Lucien. 1959. *Elements de syntaxe structurale.* Paris: Klincksieck.

Tesniere, Lucien. 2015. *Elements of Structural Syntax.* John Benjamins Publishing Company translation by timothy osborne and sylvain kahane edn.

Tucker, Gordon H. 1997. A functional lexicogrammar of adjectives. *Functions of Language* 4(2). 215–250.

Tucker, Gordon H. 1998. *The Lexicogrammar of Adjectives: A Systemic Functional Approach to Lexis.* Bloomsbury.

Turing, Allan. 1950. Computing machinery and intelligence. *Mind* 59. 433–460.

Žolkovskij, Alexander K. & Igor A. Mel'čuk. 1967. O semantičeskom sinteze. *Problemy kibernetiki* 19(?). 177–238.

Weerasinghe, Ruvan. 1994. *Probabilistic Parsing in Systemic Functional Grammar*: University of Wales College of Cardiff dissertation.

Winograd, Terry. 1972. *Understanding natural language.* Orlando, FL, USA: Academic Press, Inc. <http://linkinghub.elsevier.com/retrieve/pii/0010028572900023>.

Zeman, Daniel, Martin Popel, Milan Straka, Jan Hajic, Joakim Nivre, Filip Ginter, Juhani Luotolahti, Sampo Pyysalo, Slav Petrov, Martin Potthast, Francis M. Tyers, Elena Badmaeva, Memduh Gokirmak, Anna Nedoluzhko, Silvie Cinková, Jaroslava Hlavácová, Václava Kettnerová, Zdenka Uresová, Jenna Kanerva, Stina Ojala, Anna Missilä, Christopher D. Manning, Sebastian Schuster, Siva Reddy, Dima Taji, Nizar Habash, Herman Leung, Marie-Catherine de Marneffe, Manuela Sanguinetti, Maria Simi, Hiroshi Kanayama, Valeria de Paiva, Kira Droganova, Héctor Martínez Alonso, Çagri Çöltekin, Umut Sulubacak, Hans Uszkoreit, Vivien Macketanz, Aljoscha Burchardt, Kim Harris, Katrin Marheinecke, Georg Rehm, Tolga Kayadelen, Mohammed Attia, Ali El-Kahky, Zhuoran Yu, Emily Pitler, Saran Lertpradit, Michael Mandl, Jesse Kirchner, Hector Fernandez Alcalde, Jana Strnadová, Esha Banerjee, Ruli Manurung, Antonio Stella, Atsuko Shimada, Sookyoung Kwak, Gustavo Mendonça, Tatiana Lando, Rattima Nitisaroj & Josie Li. 2017. Conll 2017 shared task: multilingual parsing from raw text to universal dependencies. *Proceedings of the CoNLL Shared Task* 1–19.

Zhang, Niina Ning. 2010. *Coordination in syntax*. Cambridge University Press.

Zhang, Yue & Joakim Nivre. 2011. Transition-based dependency parsing with rich non-local features. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies: short papers-volume 2*, 188–193. Association for Computational Linguistics.

Zwicky, Arnold M. 1985. Heads. *Journal of Linguistics* 21. 1–30.