Chapter 32

HPSG and Dependency Grammar

Richard Hudson
University College London

HPSG assumes Phrase Structure (PS), a partonomy, in contrast with Dependency Grammar (DG), which recognises Dependency Structure (DS), with direct relations between individual words and no multi-word phrases. The chapter presents a brief history of the two approaches, showing that DG matured in the late nineteenth century, long before the influential work by Tesnière, while Phrase Structure Grammar (PSG) started somewhat later with Bloomfield's enthusiastic adoption of Wundt's ideas. Since DG embraces almost as wide a range of approaches as PSG, the rest of the chapter focuses on one version of DG, Word Grammar. The chapter argues that classical DG needs to be enriched in ways that bring it closer to PSG: each dependent actually adds an extra node to the head, but the nodes thus created form a taxonomy, not a partonomy; coordination requires strings; and in some languages the syntactic analysis needs to indicate phrase boundaries. Another proposed extension to bare DG is a separate system of relations for controlling word order, which is reminiscent of the PSG distinction between dominance and precedence. The "head-driven" part of HPSG corresponds in Word Grammar to a taxonomy of dependencies which distinguishes grammatical functions, with complex combinations similar to HPSG's re-entrance. The chapter reviews and rejects the evidence for headless phrases, and ends with the suggestion that HPSG could easily move from PS to DG.

1 Introduction

HPSG is firmly embedded, both theoretically and historically, in the phrase-structure (PS) tradition of syntactic analysis, but it also has some interesting theoretical links to the dependency-structure (DS) tradition. This is the topic of the present chapter, so after a very simple comparison of PS and DS and a glance

Richard Hudson. 2021. HPSG and Dependency Grammar. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 1403–1448. Berlin: Language Science Press. DOI: ??

at the development of these two traditions in the history of syntax, I consider a number of issues where the traditions interact.

The basis for PS analysis is the part-whole relation between smaller units (including words) and larger phrases, so the most iconic notation uses boxes (Müller 2018: 6). In contrast, the basis for DS analysis is the asymmetrical dependency relation between two words, so in this case an iconic notation inserts arrows between words. (Although the standard notation in both traditions uses trees, these are less helpful because the lines are open to different interpretations.) The two analyses of a very simple sentence are juxtaposed in Figure 1. As in HPSG attribute-value matrices (AVMs), each rectangle represents a unit of analysis.

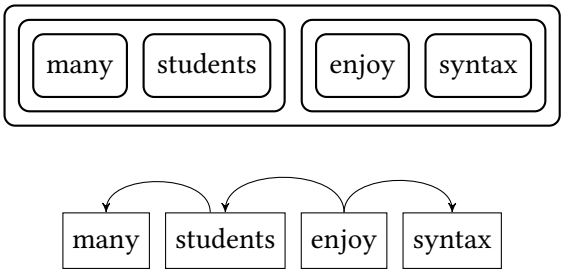

Figure 1: Phrase structure and dependency structure contrasted

In both approaches, each unit has properties such as a classification, a meaning, a form and relations to other items, but these properties may be thought of in two different ways. In PS analyses, an item contains its related items, so it also contains its other properties – hence the familiar AVMs contained within the box for each item. But in DS analyses, an item's related items are outside it, sitting alongside it in the analysis, so, for consistency, other properties may be shown as a network in which the item concerned is just one atomic node. This isn't the only possible notation, but it is the basis for the main DS theory that I shall juxtapose with HPSG, Word Grammar.

What, then, are the distinctive characteristics of the two traditions? In the following summary I use *item* to include any syntagmatic unit of analysis including morphemes, words and phrases (though this chapter will not discuss the possible role of morphemes). The following generalisations apply to classic examples of the two approaches: PS as defined by Chomsky in terms of labelled bracketed strings (Chomsky 1957), and DS as defined by Tesnière (1959; 2015). These generalisations refer to "direct relations", which are shown by single lines in standard tree notation; for example, taking a pair of words such as *big book*, they are related directly in DS, but only indirectly via a mother phrase in PS. A

phenomenon such as agreement is not a relation in this sense, but it applies to word-pairs which are identified by their relationship; so even if two sisters agree, this does not in itself constitute a direct relation between them.

1. Containment: in PS, but not in DS, if two items are directly related, one must contain the other. For instance, a PS analysis of *the book* recognises a direct relation (of dominance) between *book* and *the book*, but not between *book* and *the*, which are directly related only by linear precedence. In contrast, a DS analysis does recognise a direct relation between *book* and *the* (in addition to the linear precedence relation).

2. Continuity: therefore, in PS, but not in DS, all the items contained in a larger one must be adjacent.

3. Asymmetry: in both DS and PS, a direct relation between two items must be asymmetrical, but in DS the relation (between two words) is dependency whereas in PS it is the part-whole relation.

These generalisations imply important theoretical claims which can be tested; for instance, 2 claims that there are no discontinuous phrases, which is clearly false. On the other hand, 3 claims that there can be no exocentric or headless phrases, so DS has to consider apparent counter-examples such as NPN constructions, coordination and verbless sentences (see Sections 4.2 and 5.1 for discussion, and also Abeillé & Chaves 2021, Chapter 16 of this volume).

The contrasts in 1–3 apply without reservation to "plain vanilla" (Zwicky 1985) versions of DS and PS, but as we shall see in the history section, very few theories are plain vanilla. In particular, there are versions of HPSG that allow phrases to be discontinuous (Reape 1994; Kathol 2000; Müller 1995; 1996). Nevertheless, the fact is that HPSG evolved out of more or less pure PS, that it includes *phrase structure* in its name, and that it is never presented as a version of DS.

On the other hand, the term *head-driven* points immediately to dependency: an asymmetrical relation driven by a head word. Even if HPSG gives some constructions a headless analysis (Müller 2018: 654–666), the fact remains that it treats most constructions as headed. This chapter reviews the relations between HPSG and the very long DS tradition of grammatical analysis. The conclusion will be that in spite of its PS roots, HPSG implicitly (and sometimes even explicitly) recognises dependencies; and it may not be a coincidence that one of the main power-bases of HPSG is Germany, where the DS tradition is also at its strongest (Müller 2018: 359).

Where, then, does this discussion leave the notion of a phrase? In PS, phrases are basic units of the analysis, alongside words; but even DS recognises phrases indirectly because they are easily defined in terms of dependencies as a word plus all the words which depend, directly or indirectly, on it. Although phrases play no part in a DS analysis, it is sometimes useful to be able to refer to them informally (in much the same way that some PS grammars refer to grammatical functions informally while denying them any formal status).

Why, then, does HPSG use PS rather than DS? As far as I know, PS was simply default syntax in the circles where HPSG evolved, so the choice of PS isn't the result of a conscious decision by the founders, and I hope that this chapter will show that this is a serious question which deserves discussion.[1]Unfortunately, the historical roots and the general dominance of PS have so far discouraged discussion of this fundamental question.

HPSG is a theoretical package where PS is linked intimately to a collection of other assumptions; and the same is true for any theory which includes DS, including my own Word Grammar (Hudson 1984; 1990; 1998; 2007; 2010; Gisborne 2010; 2020; Duran-Eppler 2010; Traugott & Trousdale 2013). Among the other assumptions of HPSG I find welcome similarities, not least the use of default inheritance in some versions of the theory. I shall argue below that inheritance offers a novel solution to one of the outstanding challenges for the dependency tradition.

The next section sets the historical scene. This is important because it's all too easy for students to get the impression (mentioned above) that PS is just default syntax, and maybe even the same as "traditional grammar". We shall see that grammar has a very long and rather complicated history in which the default is actually DS rather than PS. Later sections then address particular issues shared by HPSG and the dependency tradition.

2 Dependency and constituency in the history of syntax

The relevant history of syntax starts more than two thousand years ago in Greece. (Indian syntax may have started even earlier, but it is hardly relevant because it had so little impact on the European tradition.) Greek and Roman grammarians

[1]Indeed, I once wrote a paper (which was never published) called "Taking the PS out of HPSG" – a title I was proud of until I noticed that PS was open to misreading, not least as "Pollard and Sag". Carl and Ivan took it well, and I think Carl may even have entertained the possibility that I might be right – possibly because he had previously espoused a theory called "Head Grammar" (HG).

focused on the morphosyntactic properties of individual words, but since these languages included a rich case system, they were aware of the syntactic effects of verbs and prepositions governing particular cases. However, this didn't lead them to think about syntactic relations, as such; precisely because of the case distinctions, they could easily distinguish a verb's dependents in terms of their cases: "its nominative", "its accusative" and so on (Robins 1967: 29). Both the selecting verb or preposition and the item carrying the case inflection were single words, so the Latin grammar of Priscian, written about 500 AD and still in use a thousand years later, recognised no units larger than the word: "his model of syntax was word-based – a dependency model rather than a constituency model" (Law 2003: 91). However, it was a dependency model without the notion of dependency as a relation between words.

The dependency relation, as such, seems to have been first identified by the Arabic grammarian Sibawayh in the eighth century (Owens 1988; Kouloughli 1999). However, it is hard to rule out the possibility of influence from the then-flourishing Paninian tradition in India, and in any case it doesn't seem to have had any more influence on the European tradition than did Panini's syntax, so it is probably irrelevant.

In Europe, grammar teaching in schools was based on parsing (in its original sense), an activity which was formalised in the ninth century (Luhtala 1994). The activity of parsing was a sophisticated test of grammatical understanding which earned the central place in school work that it held for centuries – in fact, right up to the 1950s (when I myself did parsing at school) and maybe beyond. In HPSG terms, school children learned a standard list of attributes for words of different classes, and in parsing a particular word in a sentence, their task was to provide the values for its attributes, including its grammatical function (which would explain its case). In the early centuries the language was Latin, but more recently it was the vernacular (in my case, English).

Alongside these purely grammatical analyses, the Ancient World had also recognised a logical one, due to Aristotle, in which the basic elements of a proposition (*logos*) are the logical subject (*onoma*) and the predicate (*rhēma*). For Aristotle a statement such as "Socrates ran" requires the recognition both of the person Socrates and of the property of running, neither of which could constitute a statement on its own (Law 2003: 30–31). By the twelfth century, grammarians started to apply a similar analysis to sentences; but in recognition of the difference between logic and grammar they replaced the logicians' *subiectum* and *praedicatum* by *suppositum* and *appositum* – though the logical terms would creep into grammar by the late eighteenth century (Law 2003: 168). This logical approach

produced the first top-down analysis in which a larger unit (the logician's proposition or the grammarian's sentence) has parts, but the parts were still single words, so *onoma* and *rhēma* can now be translated as "noun" and "verb". If the noun or verb was accompanied by other words, the older dependency analysis applied.

The result of this confusion of grammar with logic was a muddled hybrid analysis in the Latin/Greek tradition which combines a headless subject-predicate analysis with a headed analysis elsewhere, and which persists even today in some school grammars; this confusion took centuries to sort out in grammatical theory. For the subject and verb, the prestige of Aristotle and logic supported a subject-verb division of the sentence (or clause) in which the subject noun and the verb were both equally essential – a very different analysis from modern first-order logic in which the subject is just one argument (among many) which depends on the predicate. Moreover the grammatical tradition even includes a surprising number of analyses in which the subject noun is the head of the construction, ranging from the modistic grammarians of the twelfth century (Robins 1967: 83), through Henry Sweet (Sweet 1891: 17), to no less a figure than Otto Jespersen in the twentieth (Jespersen 1937), who distinguished "junction" (dependency) from "nexus" (predication) and treated the noun in both constructions as "primary".

The first grammarians to recognise a consistently dependency-based analysis for the rest of the sentence (but not for the subject and verb) seem to have been the French *encyclopédistes* of the eighteenth century (Kahane 2020), and, by the nineteenth century, much of Europe accepted a theory of sentence structure based on dependencies, but with the subject-predicate analysis as an exception – an analysis which by modern standards is muddled and complicated. Each of these units was a single word, not a phrase, and modern phrases were recognised only indirectly by allowing the subject and predicate to be expanded by dependents; so nobody ever suggested there might be such a thing as a noun phrase until the late nineteenth century. Function words such as prepositions had no proper position, being treated typically as though they were case inflections.

The invention of syntactic diagrams in the nineteenth century made the inconsistency of the hybrid analysis obvious. The first such diagram was published in a German grammar of Latin for school children (Billroth 1832), and the nineteenth century saw a proliferation of diagramming systems, including the famous Reed-Kellogg diagrams which are still taught (under the simple name "diagramming") in some American schools (Reed & Kellogg 1877); indeed, there is a website which generates such diagrams, one of which is reproduced in Figure 2.[2] The significant

[2]See a small selection of diagramming systems at http://dickhudson.com/sentence-

feature of this diagram is the special treatment given to the relation between the subject and predicate (with the verb *are* sitting uncomfortably between the two), with all the other words in the sentence linked by more or less straightforward dependencies. (The geometry of these diagrams also distinguishes grammatical functions.)

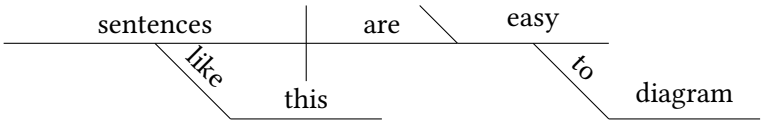

Figure 2: Reed-Kellogg diagram by Sentence Diagrammer

One particularly interesting (and relevant) fact about Reed and Kellogg is that they offer an analysis of *that old wooden house* in which each modifier creates a new unit to which the next modifier applies: *wooden house*, then *old wooden house* (Percival 1976: 18) – a clear hint at more modern structures (including the ones proposed in Section 4.1), albeit one that sits uncomfortably with plain-vanilla dependency structure.

However, even in the nineteenth century, there were grammarians who questioned the hybrid tradition which combined the subject-predicate distinction with dependencies. Rather remarkably, three different grammarians seem to have independently reached the same conclusion at roughly the same time: hybrid structures can be replaced by a homogeneous structure if we take the finite verb as the root of the whole sentence, with the subject as one of its dependents. This idea seems to have been first proposed in print in 1873 by the Hungarian Sámuel Brassai (Imrényi 2013; Imrényi & Vladár 2020); in 1877 by the Russian Aleksej Dmitrievsky (Sériot 2004); and in 1884 by the German Franz Kern (Kern 1884). Both Brassai and Kern used diagrams to present their analyses, and used precisely the same tree structures which Lucien Tesnière in France called "stemmas" nearly fifty years later (Tesnière 1959; 2015). The diagrams have both been redrawn here as Figures 3 and 4.

Brassai's proposal is contained in a school grammar of Latin, so the example is also from Latin – an extraordinarily complex sentence which certainly merits a diagram because the word order obscures the grammatical relations, which can be reconstructed only by paying attention to the morphosyntax. For example, *flentem* and *flens* both mean 'crying', but their distinct case marking links them

diagramming/ (last access October 2018), and the website is Sentence Diagrammer by 1aiway at https://download.cnet.com/windows/1aiway/3260-20_4-10725918-1.html (last access January 2019).

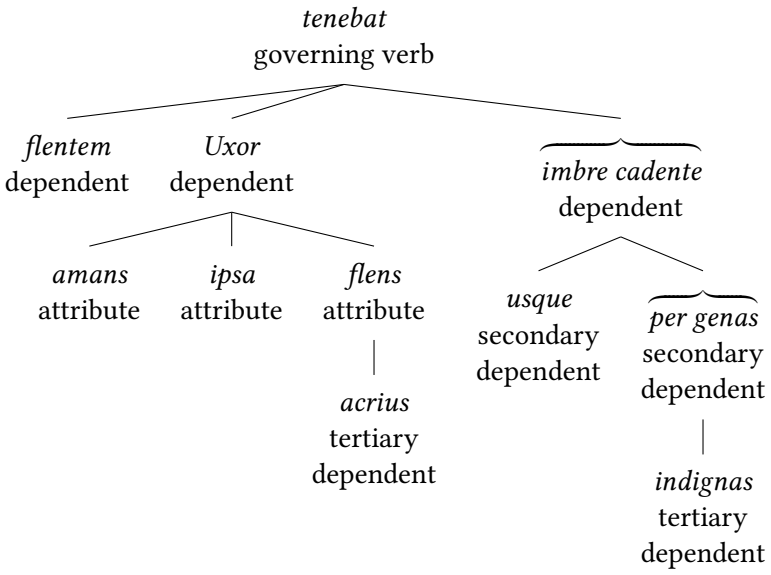

Figure 3: A verb-rooted tree published in 1873 by Brassai, quoted from Imrényi & Vladár (2020)

to different nouns, so the nominative *flens* can modify nominative *uxor* (woman), while the accusative *flentem* defines a distinct individual glossed as 'the crying one'.

(1) Uxor am-ans fl-ent-em fl-ens
wife.F.SG.NOM love-PTCP.F.SG.NOM cry-PTCP-M.SG.ACC cry-PTCP.F.SG.NOM
acr-ius ips-a ten-eb-at, imbr-e per
bitterly-more self-F.SG.NOM hug-PST-3.SG shower-M.SG.ABL on
in-dign-as usque cad-ent-e gen-as.
un-becoming-F.PL.ACC continuously fall-PTCP-M.SG.ABL cheeks-F.PL.ACC
'The wife, herself even more bitterly crying, was hugging the crying one, while a shower [of tears] was falling on her unbecoming cheeks [i.e. cheeks to which tears are unbecoming].'

Brassai's diagram, including grammatical functions as translated by the authors (Imrényi & Vladár 2020), is in Figure 3. The awkward horizontal braces should not be seen as a nod in the direction of classical PS, given that the bracketed words are not even adjacent in the sentence analysed. Kern's tree in Figure 4,

on the other hand, is for the German sentence in (2).

(2) Ein-e stolz-e Krähe schmück-t-e sich
a-F.SG.NOM proud-F.SG.NOM crow.F.SG.NOM decorate-PST-3.SG self.ACC
mit d-en aus-gefall-en-en Feder-n d-er
with the-F.PL.DAT out-fall-PTCP-F.PL.DAT feather-F.PL.DAT the-M.PL.GEN
Pfau-en.
peacock-M.PL.GEN
'A proud crow decorated himself with the dropped feathers of the peacocks.'

Once again, the original diagram includes function terms which are translated in this diagram into English.

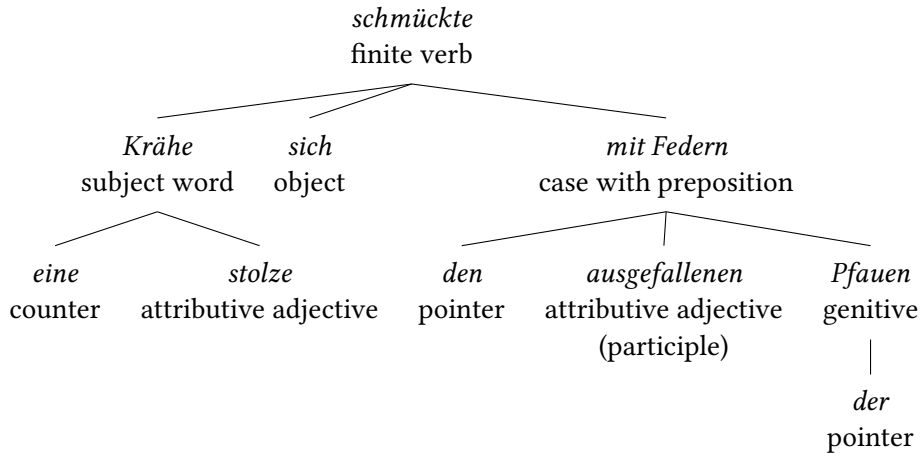

Figure 4: A verb-rooted tree from Kern (1884)

Once again the analysis gives up on prepositions, treating *mit Federn* ('with feathers') as a single word, but Figure 4 is an impressive attempt at a coherent analysis which would have provided an excellent foundation for the explosion of syntax in the next century. According to the classic history of dependency grammar, in this approach,

> [...] the sentence is not a basic grammatical unit, but merely results from combinations of words, and therefore [...] the only truly basic grammatical unit is the word. A language, viewed from this perspective, is a collection

> of words and ways of using them in word-groups, i.e., expressions of varying length. (Percival 1976: 21)

But the vagaries of intellectual history and geography worked against this intellectual breakthrough. When Leonard Bloomfield was looking for a theoretical basis for syntax, he could have built on what he had learned at school:

> […] we do not know and may never know what system of grammatical analysis Bloomfield was exposed to as a schoolboy, but it is clear that some of the basic conceptual and terminological ingredients of the system that he was to present in his 1914 and 1933 books were already in use in school grammars of English current in the United States in the nineteenth century. Above all, the notion of sentence "analysis", whether diagrammable or not, had been applied in those grammars. (Percival 1976: 18)

And when he visited Germany in 1913–1914, he might have learned about Kern's ideas, which were already influential there. But instead, he adopted the syntax of the German psychologist Wilhelm Wundt. Wundt's theory applied to meaning rather than syntax, and was based on a single idea: that every idea consists of a subject and a predicate. For example, a phrase meaning "a sincerely thinking person" has two parts: *a person* and *thinks sincerely*; and the latter breaks down, regardless of the grammar, into the noun *thought* and *is sincere* (Percival 1976).

For all its reliance on logic rather than grammar, the analysis is a clear precursor to neo-Bloomfieldian trees: it recognises a single consistent part-whole relationship (a partonomy) which applies recursively. This, then, is the beginning of the PS tradition: an analysis based purely on meaning as filtered through a speculative theory of cognition – an unpromising start for a theory of syntax. However, Bloomfield's school experience presumably explains why he combined Wundt's partonomies with the hybrid structures of Reed-Kellogg diagrams in his classification of structures as endocentric (headed) or exocentric (headless). For him, exocentric constructions include the subject-predicate structure and preposition phrases, both of which were problematic in sentence analysis at school. Consequently, his Immediate Constituent Analysis (ICA) perpetuated the old hybrid mixture of headed and headless structures.

The DS elements of ICA are important in evaluating the history of PS, because they contradict the standard view of history expressed here:

> Within the Bloomfieldian tradition, there was a fair degree of consensus regarding the application of syntactic methods as well as about the anal-

> yses associated with different classes of constructions. Some of the general features of IC analyses find an obvious reflex in subsequent models of analysis. Foremost among these is the idea that structure involves a part–whole relation between elements and a larger superordinate unit, rather than an asymmetrical dependency relation between elements at the same level. (Blevins & Sag 2013: 202–203)

This quotation implies, wrongly, that ICA rejected DS altogether.

What is most noticeable about the story so far is that, even in the 1950s, we still haven't seen an example of pure phrase structure. Every theory visited so far has recognised dependency relations in at least some constructions. Even Bloomfieldian ICA had a place for dependencies, though it introduced the idea that dependents might be phrases rather than single words and it rejected the traditional grammatical functions such as subject and object. Reacting against the latter gap, and presumably remembering their schoolroom training, some linguists developed syntactic theories which were based on constituent structure but which did have a place for grammatical functions, though not for dependency as such. The most famous of these theories are Tagmemics (Pike 1954) and Systemic Functional Grammar (Halliday 1961; 1967). However, in spite of its very doubtful parentage and its very brief history, by the 1950s virtually every linguist in America seemed to accept without question the idea that syntactic structure was a partonomy.

This is the world in which Noam Chomsky introduced phrase structure, which he presented as a formalisation of ICA, arguing that "customarily, linguistic description on the syntactic level is formulated in terms of constituent analysis (parsing)" (Chomsky 1957: 26). But such analysis was only "customary" among the Bloomfieldians, and was certainly not part of the classroom activity of parsing (Matthews 1993: 147).

Chomsky's phrase structure continued the drive towards homogeneity which had led to most of the developments in syntactic theory since the early nineteenth century. Unfortunately, Chomsky dismissed both dependencies and grammatical functions as irrelevant clutter, leaving nothing but part-whole relations, category-labels, continuity and sequential order.

Rather remarkably, the theory of phrase structure implied the (psychologically implausible) claim that sideways relations such as dependencies between individual words are impossible in a syntactic tree – or at least that, even if they are psychologically possible, they can (and should) be ignored in a formal model. Less surprisingly, having defined PS in this way, Chomsky could easily prove that it was inadequate and needed to be greatly expanded beyond the plain-vanilla

version. His solution was the introduction of transformations, but it was only thirteen years before he also recognised the need for some recognition of head-dependent asymmetries in X-bar theory (Chomsky 1970). At the same time, others had objected to transformations and started to develop other ways of making PS adequate. One idea was to include grammatical functions; this idea was developed variously in LFG (Bresnan 1978; 2001), Relational Grammar (Perlmutter & Postal 1983; Blake 1990) and Functional Grammar (Dik 1989; Siewierska 1991). Another way forward was to greatly enrich the categories (Harman 1963) as in GPSG (Gazdar et al. 1985) and HPSG (Pollard & Sag 1994).

Meanwhile, the European ideas about syntactic structure culminating in Kern's tree diagram developed rather more slowly. Lucien Tesnière in France wrote the first full theoretical discussion of DS in 1939, but it was not published till 1959 (Tesnière 1959; 2015), complete with stemmas looking like the diagrams produced seventy years earlier by Brassai and Kern. Somewhat later, these ideas were built into theoretical packages in which DS was bundled with various other assumptions about levels and abstractness. Here the leading players were from Eastern Europe, where DS flourished: the Russian Igor Mel'čuk (Mel'čuk 1988), who combined DS with multiple analytical levels, and the Czech linguists Petr Sgall, Eva Hajičová and Jarmila Panevova (Sgall et al. 1986), who included information structure. My own theory Word Grammar (developed, exceptionally, in the UK), also stems from the 1980s (Hudson 1984; 1990; Sugayama 2002; Hudson 2007; Gisborne 2008; Rosta 2008; Gisborne 2010; Hudson 2010; Gisborne 2011; Duran-Eppler 2010; Traugott & Trousdale 2013; Duran-Eppler et al. 2017; Hudson 2016; 2017; 2018; Gisborne 2019). This is the theory which I compare below with HPSG, but it is important to remember that other DS theories would give very different answers to some of the questions that I raise.

DS certainly has a low profile in theoretical linguistics, and especially so in anglophone countries, but there is an area of linguistics where its profile is much higher (and which is of particular interest to the HPSG community): natural-language processing (Kübler et al. 2009). For example:

- the Wikipedia entry for "Treebank" classifies 50 of its 101 treebanks as using DS.[3]
- The "Universal dependencies" website lists more than 100 dependency-based treebanks for 60 languages.[4]

[3] https://en.wikipedia.org/wiki/Treebank (last access January 2019).
[4] https://universaldependencies.org/ (last access January 2019).

- Google's n-gram facility allows searches based on dependencies.[5]
- The Stanford Parser (Chen & Manning 2014; Marneffe et al. 2014) uses DS.[6]

The attraction of DS in NLP is that the only units of analysis are words, so at least these units are given in the raw data and the overall analysis can immediately be broken down into a much simpler analysis for each word. This is as true for a linguist building a treebank as it was for a school teacher teaching children to parse words in a grammar lesson. Of course, as we all know, the analysis actually demands a global view of the entire sentence, but at least in simple examples a bottom-up word-based view will also give the right result.

To summarise this historical survey, PS is a recent arrival, and is not yet a hundred years old. Previous syntacticians had never considered the possibility of basing syntactic analysis on a partonomy. Instead, it had seemed obvious that syntax was literally about how words (not phrases) combined with one another.

3 HPSG and Word Grammar

The rest of this chapter considers a number of crucial issues that differentiate PS from DS by focusing specifically on how they distinguish two particular manifestations of these traditions, HPSG and Word Grammar (WG). The main question is, of course, how strong the evidence is for the PS basis of HPSG, and how easily this basis could be replaced by DS.

The comparison requires some understanding of WG, so what follows is a brief tutorial on the parts of the theory which will be relevant in the following discussion. Like HPSG, WG combines claims about syntactic relations with a number of other assumptions; but for WG, the main assumption is the Cognitive Principle:

The Cognitive Principle Language uses the same general cognitive processes and resources as general cognition, and has access to all of them.

This principle is of course merely a hypothesis which may turn out to be wrong, but so far it seems correct (Müller 2018: 494), and it is more compatible with HPSG than with the innatist ideas underlying Chomskyan linguistics (Berwick et al. 2013). In WG, it plays an important part because it determines other parts of the theory.

[5] https://books.google.com/ngrams/info and search for "dependency" (last access January 2019).
[6] https://nlp.stanford.edu/software/stanford-dependencies.shtml (last access January 2019).

On the one hand, cognitive psychologists tend to see knowledge as a network of related concepts (Reisberg 2007: 252), so WG also assumes that the whole of language, including grammar, is a conceptual network (Hudson 1984: 1; 2007: 1). One of the consequences is that the AVMs of HPSG are presented instead as labelled network links; for example, we can compare the elementary example in (3) of the HPSG lexical item for a German noun (Müller 2018: 264) with an exact translation using WG notation.

(3) AVM for the German noun *Grammatik*:

$$\begin{bmatrix} \textit{word} \\ \textsc{phonology} & \langle \textit{Grammatik} \rangle \\ \textsc{syntax-semantics} \ \ldots & \begin{bmatrix} \textit{local} \\ \textsc{category} & \begin{bmatrix} \textit{category} \\ \textsc{head} & \begin{bmatrix} \textit{noun} \\ \textsc{case} \ \boxed{1} \end{bmatrix} \\ \textsc{spr} & \langle \textsc{Det}[\textsc{case} \ \boxed{1}] \rangle \\ \ldots \end{bmatrix} \\ \textsc{content} & \ldots \begin{bmatrix} \textit{grammatik} \\ \textsc{inst} \ X \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

HPSG regards AVMs as equivalent to networks, so translating this AVM into network notation is straightforward; however, it is visually complicated, so I take it in two steps. First I introduce the basic notation in Figure 5: a small triangle showing that the lexeme GRAMMATIK "isa" word, and a headed arrow representing a labelled attribute (here, "phonology") and pointing to its value. The names of entities and attributes are enclosed in rectangles and ellipses respectively.

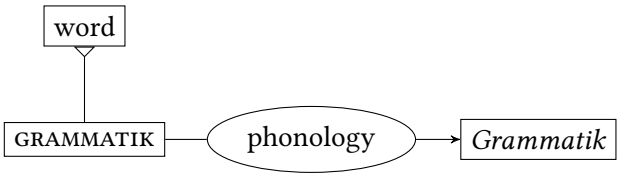

Figure 5: The German noun *Grammatik* 'grammar' in a WG network

The rest of the AVM translates quite smoothly (ignoring the list for SPR), giving Figure 6, though an actual WG analysis would be rather different in ways that are irrelevant here.

The other difference based on cognitive psychology between HPSG and WG

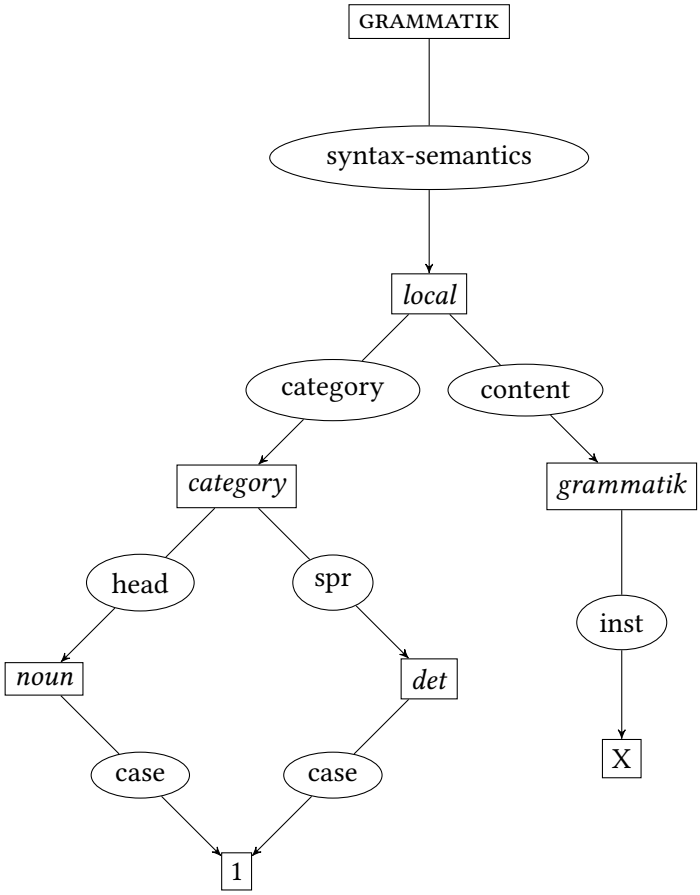

Figure 6: The German noun *Grammatik* 'grammar' in a WG network

is that many cognitive psychologists argue that concepts are built around prototypes (Rosch 1973; Taylor 1995), clear cases with a periphery of exceptional cases. This claim implies the logic of default inheritance (Briscoe et al. 1993), which is popular in AI, though less so in logic. In HPSG, default inheritance is accepted by some (Lascarides & Copestake 1999) but not by others (Müller 2018: 403), whereas in WG it plays a fundamental role, as I show in Section 4.1 below. WG uses the isa relation to carry default inheritance, and avoids the problems of non-monotonic inheritance by restricting inheritance to node-creation (Hudson 2018: 18). Once again, the difference is highly relevant to the comparison of PS and DS because one of the basic questions is whether syntactic structures in-

volve partonomies (based on whole:part relations) or taxonomies (based on the isa relation). (I argue in Section 4.1 that taxonomies exist within the structure of a sentence thanks to isa relations between tokens and sub-tokens.)

Default inheritance leads to an interesting comparison of the ways in which the two theories treat attributes. On the one hand, they both recognise a taxonomy in which some attributes are grouped together as similar; for example, the HPSG analysis in (3) classifies the attributes CATEGORY and CONTENT as LOCAL, and within CATEGORY it distinguishes the HEAD and SPECIFIER attributes. In WG, attributes are called relations, and they too form a taxonomy. The simplest examples to present are the traditional grammatical functions, which are all subtypes of "dependent"; for example, "object" isa "complement", which isa "valent", which isa "dependent", as shown in Figure 7 (which begs a number of analytical questions such as the status of depictive predicatives, which are not complements).

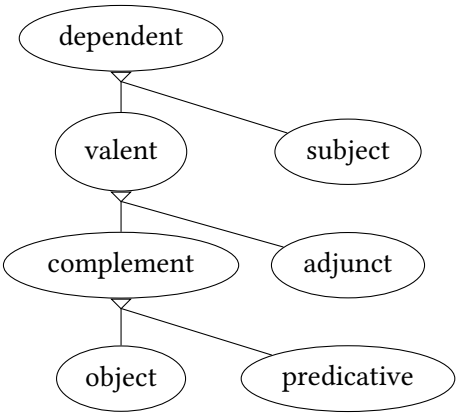

Figure 7: A WG taxonomy of grammatical functions

In spite of the differences in the categories recognised, the formal similarity is striking. On the other hand, there is also an important formal difference in the roles played by these taxonomies. In spite of interesting work on default inheritance (Lascarides & Copestake 1999), most versions of HPSG allow generalisations but not exceptions ("If one formulates a restriction on a supertype, this automatically affects all of its subtypes"; Müller 2018: 275), whereas in WG the usual logic of default inheritance applies so exceptions are possible. These are easy to illustrate from word order, which (as explained in Section 4.4) is normally inherited from dependencies: a verb's subject normally precedes it, but an inverted subject (the subject of an inverted auxiliary verb, as in *did he*) follows

it.

Another reason for discussing default inheritance and the isa relation is to explain that WG, just like HPSG, is a constraint-based theory. In HPSG, a sentence is grammatical if it can be modelled given the structures and lexicon provided by the grammar, which are combined with each other via unification. Similarly, in WG it is grammatical if its word tokens can all be inherited from entries in the grammar (which, unlike HPSG, also includes the entire lexicon). Within the grammar, these may involve overrides, but overrides between the grammar and the word tokens imply some degree of ungrammaticality. For instance, *He slept* is grammatical because all the properties of *He* and *slept* (including their syntactic properties such as the word order that can be inherited from their grammatical function) can be inherited directly from the grammar, whereas **Slept he* is ungrammatical in that the order of words is exceptional, and the exception is not licensed by the grammar.

This completes the tutorial on WG, so we are now ready to consider the issues that distinguish HPSG from this particular version of DS. In preparation for this discussion, I return to the three distinguishing assumptions about classical PS and DS theories given earlier as 1 to 3, and repeated here:

1. Containment: in PS, but not in DS, if two items are directly related, one must contain the other.
2. Continuity: therefore, in PS, but not in DS, all the items contained in a larger one must be adjacent.
3. Asymmetry: in both DS and PS, a direct relation between two items must be asymmetrical, but in DS the relation (between two words) is dependency whereas in PS it is the part-whole relation.

These distinctions will provide the structure for the discussion:

- Containment and continuity:
  - Semantic phrasing
  - Coordination
  - Phrasal edges
  - Word order
- Asymmetry:

- Structure sharing and raising/lowering
- Headless phrases
- Complex dependency
- Grammatical functions

4 Containment and continuity (PS but not DS)

4.1 Semantic phrasing

One apparent benefit of PS is what I call "semantic phrasing" (Hudson 1990: 146–151), in which the effect of adding a dependent to a word modifies that word's meaning to produce a different meaning. For instance, the phrase *typical French house* does not mean 'house which is both typical and French', but rather 'French house which is typical (of French houses)' (Dahl 1980). In other words, even if the syntax does not need a node corresponding to the combination *French house*, the semantics does need one.

For HPSG, of course, this is not a problem, because every dependent is part of a new structure, semantic as well as syntactic (Müller 2019); so the syntactic phrase *French house* has a content which is 'French house'. But for DS theories, this is not generally possible, because there is no syntactic node other than those for individual words – so, in this example, one node for *house* and one for *French* but none for *French house*.

Fortunately for DS, there is a solution: create extra word nodes but treat them as a taxonomy, not a partonomy (Hudson 2018). To appreciate the significance of this distinction, the connection between the concepts "finger" and "hand" is a partonomy, but that between "index finger" and "finger" is a taxonomy; a finger is part of a hand, but it is not a hand, and conversely an index finger is a finger, but it is not part of a finger.

In this analysis, then, the token of *house* in *typical French house* would be factored into three distinct nodes:

- *house*: an example of the lexeme HOUSE, with the inherited meaning 'house'.
- *house+F*: the word *house* with *French* as its dependent, meaning 'French house'.
- *house+t*: the word *house+F* with *typical* as its dependent, meaning 'typical example of a French house'.

(It is important to remember that the labels are merely hints to guide the analyst, and not part of the analysis; so the last label could have been *house+t+F* without changing the analysis at all. One of the consequences of a network approach is that the only substantive elements in the analysis are the links between nodes, rather than the labels on the nodes.) These three nodes can be justified as distinct categories because each combines a syntactic fact with a semantic one: for instance, *house* doesn't simply mean 'French house', but has that meaning because it has the dependent *French*. The alternative would be to add all the dependents and all the meanings to a single word node as in earlier versions of WG (Hudson 1990: 146–151), thereby removing all the explanatory connections; this seems much less plausible psychologically. The proposed WG analysis of *typical French house* is shown in Figure 8, with the syntactic structure on the left and the semantics on the right.

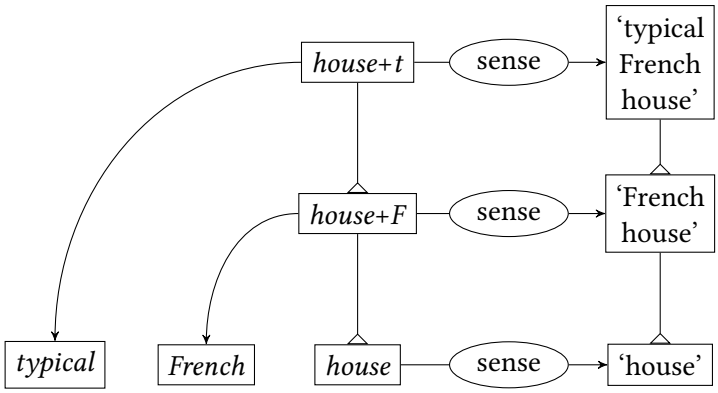

Figure 8: *typical French house* in WG

Unlike standard DG analyses (Müller 2019), the number of syntactic nodes in this analysis is the same as in an HPSG analysis, but crucially these nodes are linked by the isa relation, and not as parts to wholes – in other words, the hierarchy is a taxonomy, not a partonomy. As mentioned earlier, the logic is default inheritance, and the default semantics has isa links parallel to those in syntax; thus the meaning of *house+F* (*house* as modified by *French*) isa the meaning of *house* – in other words, a French house is a kind of house. But the default can be overridden by exceptions such as the meanings of adjectives like *fake* and *former*, so a fake diamond is not a diamond (though it looks like one) and a former soldier is no longer a soldier. The exceptional semantics is licensed by the grammar – the stored network – so the sentence is fully grammatical. All this is possible

because of the same default inheritance that allows irregular morphology and syntax.

4.2 Coordination

Another potential argument for PS, and against DS, is based on coordination: coordination is a symmetrical relationship, not a dependency, and it coordinates phrases rather than single words. For instance, in (4) the coordination clearly links the VPs *came in* to *sat down* and puts them on equal grammatical terms; and it is this equality that allows them to share the subject *Mary*.

(4) Mary came in and sat down.

But of course, in a classic DS analysis *Mary* is also attached directly to *came*, without an intervening VP node, so *came in* is not a complete syntactic item and this approach to coordination fails, so we have a prima facie case against DS. (For coordination in HPSG, see Abeillé & Chaves 2021, Chapter 16 of this volume.)

Fortunately, there is a solution: sets (Hudson 1990: 404–421). We know from the vast experimental literature (as well as from everyday experience) that the human mind is capable of representing ordered sets (strings) of words, so all we need to assume is that we can apply this ability in the case of coordination. The members of a set are all equal, so their relation is symmetrical; and the members may share properties (e.g. a person's children constitute a set united by their shared relation to that person as well as by a multitude of other shared properties). Moreover, sets may be combined into supersets, so both conjuncts such as *came in* and *sat down* and coordinations (*came in* and *sat down*) are lists. According to this analysis, then, the two lists (*came*, *in*) and (sat, down) are united by their shared subject, Mary, and combine into the coordination ((*came*, *in*) (*sat*, *down*)). The precise status of the conjunction *and* remains to be determined. The proposed analysis is shown in network notation in Figure 9.

Once again, inheritance plays a role in generating this diagram. The isa links have been omitted in Figure 9 to avoid clutter, but they are shown in Figure 10, where the extra isa links are compensated for by removing all irrelevant matter and the dependencies are numbered for convenience. In this diagram, the dependency d1 from *came* to *Mary* is the starting point, as it is established in processing during the processing of *Mary came* – long before the coordination is recognised; and the endpoint is the dependency d5 from *sat* to *Mary*, which is simply a copy of d1, so the two are linked by isa. (It will be recalled from Figure 7 that dependencies form a taxonomy, just like words and word classes, so isa links between

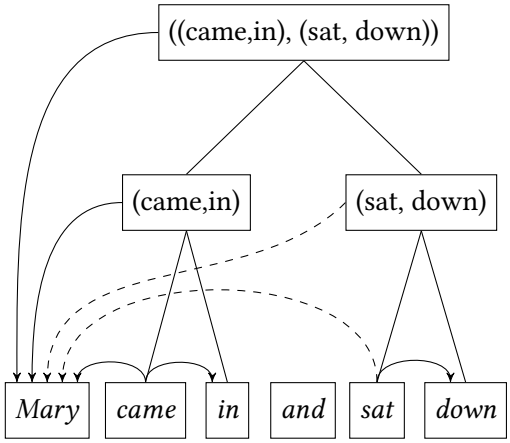

Figure 9: Coordination with sets

dependencies are legitimate.) The conjunction *and* creates the three set nodes, and general rules for sets ensure that properties – in this case, dependencies – can be shared by the two conjuncts.

It's not yet clear exactly how this happens, but one possibility is displayed in the diagram: d1 licenses d2 which licenses d3 which licenses d4 which licenses d5. Each of these licensing relations is based on isa. Whatever the mechanism, the main idea is that the members of a set can share a property; for example, we can think of a group of people sitting in a room as a set whose members share the property of sitting in the room. Similarly, the set of strings *came in* and *sat down* share the property of having *Mary* as their subject.

The proposed analysis may seem to have adopted phrases in all but name, but this is not so because the conjuncts have no grammatical classification, so coordination is not restricted to coordination of like categories. This is helpful with examples like (5) where an adjective is coordinated with an NP and a PP.

(5) Kim was intelligent, a good linguist and in the right job.

The possibility of coordinating mixed categories is a well-known challenge for PS-based analyses such as HPSG: "Ever since Sag et al. (1985), the underlying intuition was that what makes Coordination of Unlikes acceptable is that each conjunct is actually well-formed when combined individually with the shared rest" (Crysmann 2008: 61). Put somewhat more precisely, the intuition is that what coordinated items share is not their category but their function (Hudson 1990: 414). This is more accurate because simple combinability isn't enough; for

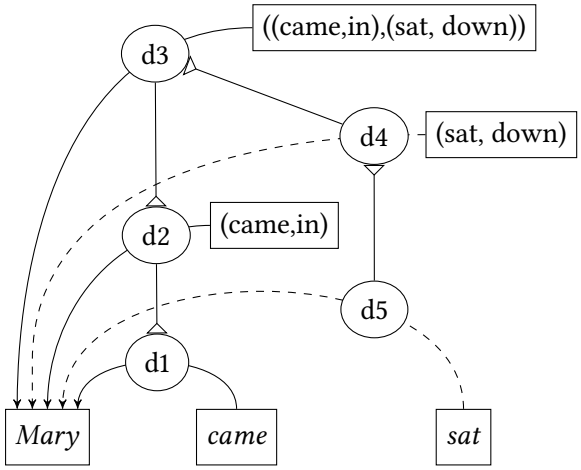

Figure 10: Coordination with inherited dependencies

instance, *we ate* can combine with an object or with an adjunct, but the functional difference prevents them from coordinating:

(6) We ate a sandwich.

(7) We ate at midday.

(8) * We ate a sandwich and at midday.

Similarly, *a linguist* can combine as dependent with many verbs, but these can only coordinate if their relation to *a linguist* is the same:

(9) She became a linguist.

(10) She met a linguist.

(11) * She became and met a linguist.

It is true that HPSG can accommodate the coordination of unlike categories by redefining categories so that they define functions rather than traditional categories; for example, if "predicative" is treated as a category, then the problem of (5) disappears because *intelligent*, *a good linguist* and *in the right job* all belong to the category "predicative". However, this solution generates as many problems as it solves. For example, why is the category "predicative" exactly equivalent to the function with the same name, whereas categories such as "noun phrase" have multiple functions? And how does this category fit into a hierarchy of categories

so as to bring together an arbitrary collection of categories which are otherwise unrelated: nominative noun phrase, adjective phrase and preposition phrase?

Moreover, since the WG analysis is based on arbitrary strings and sets rather than phrases, it easily accommodates "incomplete" conjuncts (Hudson 1990: 405; Hudson 1982) precisely because there is no expectation that strings are complete phrases. This claim is born out by examples such as (12) (meaning '… and parties for foreign girls …').

(12) We hold parties for foreign boys on Tuesdays and girls on Wednesdays.

In this example, the first conjunct is the string (*boys*, *on*, *Tuesdays*), which is not a phrase defined by dependencies; the relevant phrases are *parties for foreign boys* and *on Tuesdays*.

This sketch of a WG treatment of coordination ignores a number of important issues (raised by reviewers) such as joint interpretation (13) and special choice of pronoun forms (14).

(13) John and Mary are similar.

(14) Between you and I, she likes him.

These issues have received detailed attention in WG (Hudson 1984: Chapter 5; 1988; 1990: Chapter 14; 1995; 2010: 175–181, 304–307), but they are peripheral to this chapter.

4.3 Phrasal edges

One of the differences between PS and DS is that, at least in its classic form, PS formally recognises phrasal boundaries, and a PS tree can even be converted to a bracketed string where the phrase is represented by its boundaries. In contrast, although standard DS implies phrases (since a phrase can be defined as a word and all the words depending on it either directly or indirectly), it doesn't mark their boundaries.

This turns out to be problematic in dealing with Welsh soft mutation (Tallerman 2009). Tallerman's article is one of the few serious discussions by a PS advocate of the relative merits of PS and DS, so it deserves more consideration than space allows here. It discusses examples such as (15) and (16), where the underlined words are morphologically changed by soft mutation in comparison with their underlying forms shown in brackets.

(15) Prynodd y ddynes delyn. (telyn)
buy.PAST.3S the woman harp
'The woman bought a harp.'

(16) Gwnaeth y ddynes [werthu telyn]. (gwerthu)
do.PAST.3S the woman sell.INF harp
'The woman sold a harp.'

Soft mutation is sensitive to syntax, so although 'harp' is the object of a preceding verb in both examples, it is mutated when this verb is finite (*prynodd*) and followed by a subject, but not when the verb has no subject because it is non-finite (*werthu*). Similarly, the non-finite verb 'sell' is itself mutated in example (16) because it follows a subject, in contrast with the finite verbs which precede the subject and have no mutation.

A standard PS explanation for such facts (and many more) is the "XP Trigger Hypothesis": that soft mutation is triggered on a subject or complement (but not an adjunct) immediately after an XP boundary (Borsley et al. 2007: 226). The analysis contains two claims: that mutation affects the first word of an XP, and that it is triggered by the end of another XP. The first claim seems beyond doubt: the mutated word is simply the first word, and not necessarily the head. Examples such as (17) are conclusive.

(17) Dw i [lawn mor grac â chi]. (llawn)
be.PRES.1S I full as angry as you
'I'm just as angry as you.'

The second claim is less clearly correct; for instance, it relies on controversial assumptions about null subjects and traces in examples such as (18) and (19) (where *t* and *pro* stand for a trace and a null subject respectively, but have to be treated as full phrases for purposes of the XP Trigger Hypothesis in order to explain the mutation following them).

(18) Pwy brynodd *t* delyn? (telyn)
who buy.PAST.3S *t* harp
'Who bought a harp?'

(19) Prynodd *pro* delyn. (telyn)
buy.PAST.3S *pro* harp
'He/she bought a harp.'

But suppose both claims were true. What would this imply for DS? All it shows is that we need to be able to identify the first word in a phrase (the mutated word) and the last word in a phrase (the trigger). This is certainly not possible in WG as it stands, but the basic premise of WG is that the whole of ordinary cognition is available to language, and it's very clear that ordinary cognition allows us to recognise beginnings and endings in other domains, so why not also in language? Moreover, beginnings and endings fit well in the framework of ideas about linearisation that are introduced in the next subsection.

The Welsh data, therefore, do not show that we need phrasal nodes complete with attributes and values. Rather, edge phenomena such as Welsh mutation show that DS needs to be expanded, but not that we need the full apparatus of PS. Exactly how to adapt WG is a matter for future research, not for this chapter.

4.4 Word order

In both WG and some variants of HPSG, dominance and linearity are separated, but this separation goes much further in WG. In basic HPSG, linearisation rules apply only to sisters, and if the binary branching often assumed for languages such as German (Müller 2018: Section 10.3) reduces these to just two, the result is clearly too rigid given the freedom of ordering found in many languages. It is true that solutions are available (Müller 2018: Chapter 10), such as allowing alternative binary branchings for the same word combinations or combining binary branching with flat structures held in lists, but these solutions involve extra complexity in other parts of the theory such as additional lists. For instance, one innovation is the idea of linearisation domains (Reape 1994; Kathol 2000; Müller 1996), which allow a verb and its arguments and adjuncts to be members of the same linearisation domain and hence to be realized in any order (Müller 2018: 302). These proposals bring HPSG nearer to DS, where flat structures are inevitable and free order is the default (subject to extra order constraints).

WG takes the separation of linearity from dominance a step further by introducing two new syntactic relations dedicated to word order: "position" and "landmark", each of which points to a node in the overall network (Hudson 2018). As its name suggests, a word's landmark is the word from which it takes its position, and is normally the word on which it depends (as in the HPSG list of dependents); what holds phrases together by default is that dependents keep as close to their landmarks as possible, because a general principle bans intersecting landmark relations. Moreover, the word's "position" relative to its landmark may either be free or defined as either "before" or "after".

However, this default pattern allows exceptions, and because "position" and

"landmark" are properties, they are subject to default inheritance which allows exceptions such as raising and extraction (discussed in Section 5.2). To give an idea of the flexibility allowed by these relations, we start with the very easy English example in Figure 11, where "lm" and "psn" stand for "landmark" and "position", and "<" and ">" mean "before" and "after".

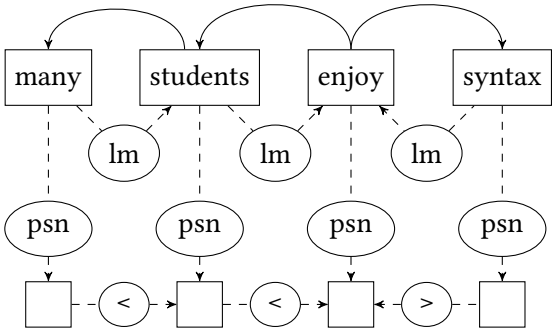

Figure 11: Basic word order in English

It could be objected that this is a lot of formal machinery for such a simple matter as word order. However, it is important to recognise that the conventional left-right ordering of writing is just a written convention, and that a mental network (which is what we are trying to model in WG) has no left-right ordering. Ordering a series of objects (such as words) is a complex mental operation, which experimental subjects often get wrong, so complex machinery is appropriate. Moreover, any syntactician knows that language offers a multiplicity of complex relations between dependency structure and word order. To take an extreme example, non-configurational languages pose problems for standard versions of HPSG (for which Bender suggests solutions) as illustrated by a Wambaya sentence, repeated here as (20) (Bender 2008; Nordlinger 1998):

(20) Ngaragana-nguja ngiy-a gujinganjanga-ni jiyawu ngabulu
grog-PROP.IV.ACC 3.SG.NM.A-PST mother-II.ERG give milk.IV.ACC
'(His) mother gave (him) milk with grog in it.'

The literal gloss shows that both 'grog' and 'milk' are marked as accusative, which is enough to allow the former to modify the latter in spite of their separation. The word order is typical of many Australian non-configurational languages: totally free within the clause except that the auxiliary verb (glossed here as 3SG.PAST) comes second (after one dependent word or phrase). Such freedom

of order is easily accommodated if landmarks are independent of dependencies: the auxiliary verb is the root of the clause's dependency structure (as in English), and also the landmark for every word that depends on it, whether directly or (crucially) indirectly. Its second position is due to a rule which requires it to precede all these words by default, but to have just one "preceder". A simplified structure for this sentence (with Wambaya words replaced by English glosses) is shown in Figure 12, with dotted arrows below the words again showing landmark and position relations. The dashed horizontal line separates this sentence structure from the grammar that generates it. In words, an auxiliary verb requires precisely one preceder, which isa descendant. "Descendant" is a transitive generalisation of "dependent", so a descendant is either a dependent or a dependent of a descendant. The preceder precedes the auxiliary verb, but all other descendants follow it.

Later sections will discuss word order, and will reinforce the claims of this subsection: that plain-vanilla versions of either PS or DS are woefully inadequate and need to be supplemented in some way.

This completes the discussion of "containment" and "continuity", the characteristics of classical PS which are missing in DS. We have seen that the continuity guaranteed by PS is also provided by default in WG by a general ban on intersecting landmark relations; but, thanks to default inheritance, exceptions abound. HPSG offers a similar degree of flexibility but using different machinery such as word-order domains (Reape 1994); see also Müller (2021), Chapter 10 of this volume. Moreover, WG offers a great deal of flexibility in other relations: for example, a word may be part of a string (as in coordination) and its phrase's edges may need to be recognised structurally (as in Welsh mutation).

5 Asymmetry and functions

This section considers the characteristics of DS which are missing from classical PS: asymmetrical relations between words and their dependents. Does syntactic theory need these notions? It's important to distinguish here between two different kinds of asymmetry that are recognised in HPSG. One is the kind which is inherent to PS and the part-whole relation, but the other is inherent to DS but an optional extra in PS: the functional asymmetry between the head and its dependents. HPSG, like most other theories of syntax, does recognise this asymmetry and indeed builds it into the name of the theory, but more recently this assumption has come under fire within the HPSG community for reasons considered below in Section 5.1.

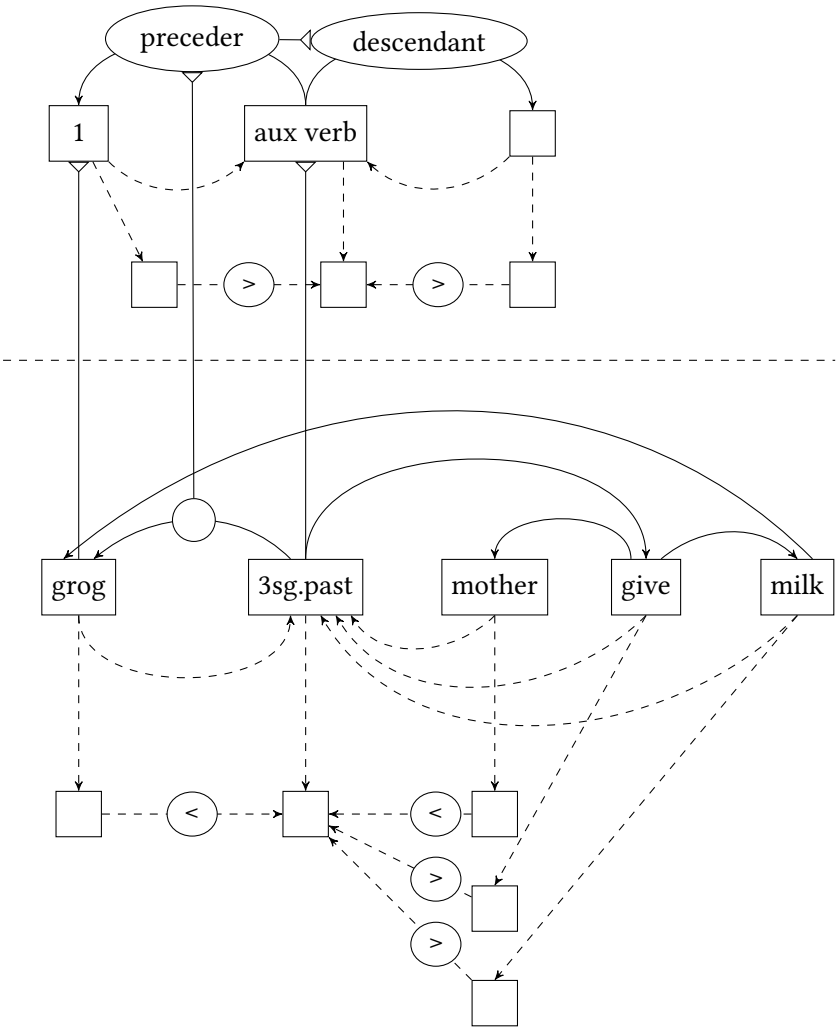

Figure 12: A non-configurational structure

But if the head/dependent distinction is important, are there any other functional distinctions between parts that ought to be explicit in the analysis? In other words, what about grammatical functions such as subject and object? As Figure 7 showed, WG recognises a taxonomy of grammatical functions which carry important information about word order (among other things), so functions are central to WG analyses. Many other versions of DS also recognise functional distinctions; for example, Tesnière distinguished actants from circumstantials, and among actants he distinguished subjects, direct objects and indirect objects (Tesnière 2015: xlvii). But the only functional distinction which is inherent to DS is the one between head and dependents, so other such distinctions are an optional extra in DS – just as they are in PS, where many theories accept them. But HPSG leaves them implicit in the order of elements in ARG-ST (like phrases in DS), so this is an issue worth raising when comparing HPSG with the DS tradition.

5.1 Headless phrases

Bloomfield assumed that phrases could be either headed (endocentric) or not (exocentric). According to WG (and other DS theories), there are no headless phrases. Admittedly, utterances may contain unstructured lists (e.g. *one two three four* ...), and quotations may be unstructured strings, as in (21), but presumably no-one would be tempted to call such strings "phrases", or at least not in the sense of phrases that a grammar should generate.

(21) He said "One, two, three, testing, testing, testing."

Such strings can be handled by the mechanism already introduced for coordination, namely ordered sets.

The WG claim, then, is that when words hang together syntactically, they form phrases which always have a head. Is this claim tenable? There are a number of potential counterexamples including (22a)–(22d):

(22) a. The rich get richer. (Müller 2018: 403)
b. The harder he works, the less he learns. (Fillmore 1986)
c. In they came, student after student. (Jackendoff 2008: 8)
d. However smart the students, a lecture needs to be clear. (Arnold & Borsley 2014)

All these examples can in fact be given a headed analysis, as I shall now explain, starting with (22a). *The rich* is allowed by *the*, which has a special subcase which allows a single adjective as its complement, meaning either "generic

people" or some contextually defined notion (such as "apples" in *the red* used when discussing apples); this is not possible with any other determiner. In the determiner-headed analysis of standard WG, this is unproblematic as the head is *the*.

The comparative correlative in (22b) is clearly a combination of a subordinate clause followed by a main clause (Culicover & Jackendoff 1999), but what are the heads of the two clauses? The obvious dependency links the first *the* with the second (hence "correlative"), so it is at least worth considering an analysis in which this dependency is the basis of the construction and, once again, the head is *the*. Figure 13 outlines a possible analysis, though it should be noted that the dependency structures are complex. The next section discusses such complexities, which are a reaction to complex functional pressures; for example, it is easy to see that the fronting of *the less* reduces the distance between the two correlatives. Of course, there is no suggestion here that this analysis applies unchanged to every translation equivalent of our comparative correlative; for instance, French uses a coordinate structure without an equivalent of *the*: *Plus … et plus …* (Abeillé & Borsley 2008).

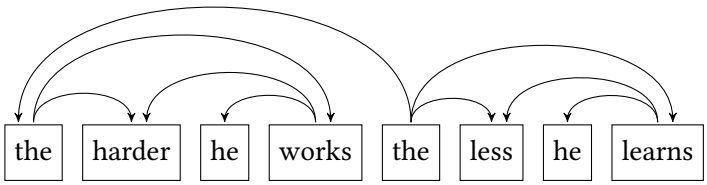

Figure 13: A WG sketch of the comparative correlative

Example (22c) is offered by Jackendoff as a clear case of headlessness, but there is an equally obvious headed analysis of *student after student* in which the structure is the same as in commonplace NPN examples like *box of matches*. The only peculiarity of Jackendoff's example is the lexical repetition, which is beyond most theories of syntax. For WG, however, the solution is easy: the second N token isa the first, which allows default inheritance. This example illustrates an idiomatic but generalisable version of the NPN pattern in which the second N isa the first and the meaning is special; as expected, the pattern is recursive. The grammatical subnetwork needed to generate the syntactic structure for such examples is shown (with solid lines) in Figure 14; the semantics is harder and needs more research. What this diagram shows is that there is a subclass of nouns called here "noun_{npn}", which is special in having as its complement a preposition with the

special property of having another copy of the same noun_{npn} as its complement. The whole construction is potentially recursive because the copy itself inherits the possibility of a preposition complement, but the recursion is limited by the fact that this complement is optional (shown as "0,1" inside the box, meaning that its quantity is either 0 (absent) or 1 (present)). Because the second noun isa the first, if it has a prepositional complement this is also a copy of the first preposition – hence *student after student after student*, whose structure is shown in Figure 14 with dashed lines.

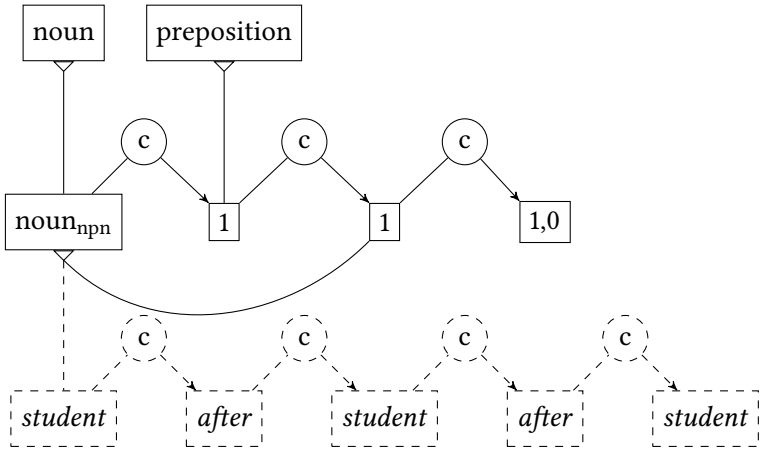

Figure 14: The NPN construction in Word Grammar

The "exhaustive conditional" or "unconditional" in (22d) clearly has two parts: *however smart* and *the students*, but which is the head? A verb could be added, giving *however smart the students are*, so if we assumed a covert verb, that would provide a head, but without a verb it is unclear – and indeed this is precisely the kind of subject-predicate structure that stood in the way of dependency analysis for nearly two thousand years.

However, there are good reasons for rejecting covert verbs in general. For instance, in Arabic a predicate adjective or nominal is in different cases according to whether "be" is overt: accusative when it is overt, nominative when it is covert. Moreover, the word order is different in the two constructions: the verb normally precedes the subject, but the verbless predicate follows it. In Arabic, therefore, a covert verb would simply complicate the analysis; but if an analysis without a covert verb is possible for Arabic, it is also possible in English.

Moreover, even English offers an easy alternative to the covert verb based on the structure where the verb BE is overt. It is reasonably uncontroversial to

assume a raising analysis for examples such as (23a) and (23b), so (23c) invites a similar analysis (Müller 2012).

(23) a. He keeps talking.
b. He is talking.
c. He is cold.

But a raising analysis implies a headed structure for *he ... cold* in which *he* depends (as subject) on *cold*. Given this analysis, the same must be true even where there is no verb, as in example (22d)'s *however smart the students* or so-called "Mad-Magazine sentences" like (24) (Lambrecht 1990).[7]

(24) What, him smart? You're joking!

Comfortingly, the facts of exhaustive conditionals support this analysis because the subject is optional, confirming that the predicate is head:

(25) However smart, nobody succeeds without a lot of effort.

In short, where there is just a subject and a predicate, without a verb, then the predicate is the head.

Clearly it is impossible to prove the non-existence of headless phrases, but the examples considered have been offered as plausible examples, so if even they allow a well-motivated headed analysis, it seems reasonable to hypothesise that all phrases have heads.

5.2 Complex dependency

The differences between HPSG and WG raise another question concerning the geometry of sentence structure, because the possibilities offered by the part-whole relations of HPSG are more limited than those offered by the word-word dependencies of WG. How complex can dependencies be? Is there a theoretical limit such that some geometrical patterns can be ruled out as impossible? Two particular questions arise:

[7] A reviewer asks what excludes alternatives such as *He smart?* and *Him smart.* (i.e. as a statement). The former is grammatically impossible because *he* is possible only as the subject of a tensed verb, but presumably the latter is excluded by the pragmatic constraints on the "Mad-magazine" construction.

1. Can a word depend on more than one other word? This is of course precisely what structure sharing allows, but this only allows "raising" or "lowering" within a single chain of dependencies. Is any other kind of "double dependency" possible?

2. Is mutual dependency possible?

The answer to both questions is yes for WG, but is less clear for HPSG. Consider the dependency structure for an example such as (26).

(26) I wonder who came.

In a dependency analysis, the only available units are words, so the clause *who came* has no status in the analysis and is represented by its head. In WG, this is *who*, because this is the word that links *came* to the rest of the sentence.

Of interest in (26) are three dependencies:

1. *who* depends on *wonder* because *wonder* needs an interrogative complement – i.e. an interrogative word such as *who* or *whether*; so *who* is the object of *wonder*.

2. *who* also depends on *came*, because it is the subject of *came*.

3. *came* depends on *who*, because interrogative pronouns allow a following finite verb (or, for most but not all pronouns, an infinitive, as in *I wonder who to invite*). Since this is both selected by the pronoun and optional (as in *I wonder who*), it must be the pronoun's complement, so *came* is the complement of *who*.

Given the assumptions of DS, and of WG in particular, each of these dependencies is quite obvious and uncontroversial when considered in isolation. The problem, of course, is that they combine in an unexpectedly complicated way; in fact, this one example illustrates both the complex conditions defined above: *who* depends on two words which are not otherwise syntactically connected (*wonder* and *came*), and *who* and *came* are mutually dependent. A WG analysis of the relevant dependencies is sketched in Figure 15 (where "s" and "c" stand for "subject" and "complement").

A similar analysis applies to relative clauses. For instance, in (27), the relative pronoun *who* depends on the antecedent *man* as an adjunct and on *called* as its subject, while the "relative verb" *called* depends on *who* as its obligatory complement.

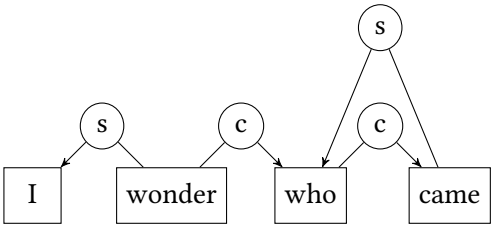

Figure 15: Complex dependencies in a relative clause

(27) I knew the man who called.

Pied-piping presents well-known challenges. Take, for example, (28) (Pollard & Sag 1994: 212).

(28) Here's the minister [[in [the middle [of [whose sermon]]]] the dog barked]

According to WG, *whose* (which as a determiner is head of the phrase *whose sermon*) is both an adjunct of its antecedent *minister* and also the head of the relative verb *barked*, just as in the simpler example. The challenge is to explain the word order: how can *whose* have dependency links to both *minister* and *barked* when it is surrounded, on both sides, by words on which it depends? Normally, this would be impossible, but pied-piping is special. The WG analysis (Hudson 2018) locates the peculiarities of pied-piping entirely in the word order, invoking a special relation "pipee" which transfers the expected positional properties of the relative pronoun (the "piper") up the dependency chain – in this case, to the preposition *in*.

And so we finish this review of complex dependencies by answering the question that exercised the minds of the Arabic grammarians in the Abbasid Caliphate: is mutual dependency possible? The arrow notation of WG allows grammars to generate the relevant structures, so the answer is yes, and HPSG can presumably achieve the same effect by means of re-entrance; so this conclusion reflects another example of theoretical convergence.

5.3 Grammatical functions

As I have already explained, more or less traditional grammatical functions such as subject and adjunct play a central part in WG, and more generally, they are highly compatible with any version of DS, because they are all sub-divisions of the basic function "dependent". This being so, we can define a taxonomy of functions such as the one in Figure 7, which is developed in Figure 16 to accommodate

an example of the very specific functions which are needed in any complete grammar: the second complement of *from*, as in *from London to Edinburgh*, which may be unique to this particular preposition.

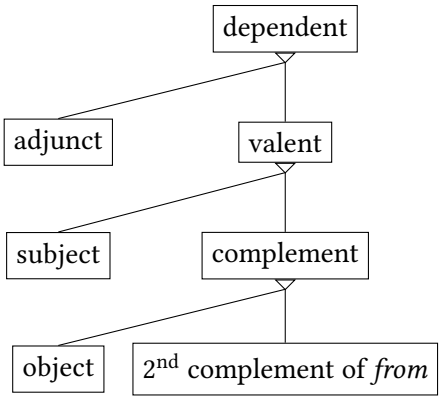

Figure 16: A taxonomy of grammatical functions

HPSG also recognises a taxonomy of functions by means of three lists attached to any head word:

SPR: the word's specifier, i.e. for a noun its determiner and (in some versions of HPSG) for a verb its subject

COMPS: its complements

ARG-ST: its specifier and its complements, i.e. in WG terms, its valents.

The third list concatenates the first two, so the same analysis could be achieved in WG by a taxonomy in which SPR and COMPS both isa ARG-ST. However, there are also two important differences: in HPSG, adjuncts have a different status from other dependents, and these three general categories are lists.

Adjuncts are treated differently in the two theories. In WG, they are dependents, and located in the same taxonomy as valents; so in HPSG terms they would be listed among the head word's attributes, along with the other dependents but differentiated by not being licensed by the head. But HPSG reverses this relationship by treating the head as a MOD ("modified") of the adjunct. For example, in (29) *she* and *it* are listed in the ARG-ST of *ate* but *quickly* is not mentioned in the AVM of *ate*; instead, *ate* is listed as MOD of *quickly*.

(29) She ate it quickly.

This distinction, inherited from Categorial Grammar, correctly reflects the facts of government: *ate* governs *she* and *it*, but not *quickly*. It also reflects one possible analysis of the semantics, in which *she* and *it* provide built-in arguments of the predicate "eat", while *quickly* provides another predicate "quick", of which the whole proposition *eat*(*she*, *it*) is the argument. Other semantic analyses are of course possible, including one in which "manner" is an optional argument; but the proposed analysis is consistent with the assumptions of HPSG.

On the other hand, HPSG also recognises a HEAD constituent in every construction, and in the construction which includes *quickly*, the latter is not the head. So what unifies arguments and adjuncts is the fact of not being heads. In contrast, DS theories (including WG) agree in recognising adjuncts as dependents, so arguments and adjuncts are unified by this category, which is missing from most versions of HPSG, though not from all (Bouma et al. 2001). The DS analysis follows from the assumption that dependency isn't just about government, nor is it tied to a logical analysis based on predicates and arguments. At least in WG, the basic characteristic of a dependent is that it modifies the meaning of the head word, so that the resultant meaning is (typically) a hyponym of the head's unmodified meaning. Given this characterisation, adjuncts are core dependents; for instance *big book* is a hyponym of *book* (i.e. "big book" isa "book"), and *she ate it quickly* is a hyponym of *she ate it.* The same characterisation also applies to arguments: *ate it* is a hyponym of *ate*, and *she ate it* is a hyponym of *ate it.* (Admittedly hyponymy is merely the default, and as explained in Section 4.1 it may be overridden by the details of particular adjuncts such as *fake* as in *fake diamonds*; but exceptions are to be expected.)

Does the absence in HPSG of a unifying category "dependent" matter? So long as HEAD is available, we can express word-order generalisations for head-final and head-initial languages, and maybe also for "head-medial" languages such as English (Hudson 2010: 172). At least in these languages, adjuncts and arguments follow the same word-order rules, but although it is convenient to have a single cover term "dependent" for them, it is probably not essential. So maybe the presence of HEAD removes the need for its complement, DEPENDENT.

The other difference between HPSG and WG lies in the way in which the finer distinctions among complements are made. In HPSG they are shown by the ordering of elements in a list, whereas WG distinguishes them as further subcategories in a taxonomy. For example, in HPSG the direct object is identified as the second NP in the ARG-ST list, but in WG it is a sub-category of "complement" in the taxonomy of Figure 16. In this case, each approach seems to offer something which is missing from the other.

On the one hand, the ordered lists of HPSG reflect the attractive ranking of dependents offered by Relational Grammar (Perlmutter & Postal 1983; Blake 1990) in which arguments are numbered from 1 to 3 and can be "promoted" or "demoted" on this scale. The scale had subjects at the top and remote adjuncts at the bottom, and appeared to explain a host of facts from the existence of argument-changing alternations such as passivisation (Levin 1993) to the relative accessibility of different dependents to relativisation (Keenan & Comrie 1977). An ordered list, as in ARG-ST, looks like a natural way to present this ranking of dependents.

On the other hand, the taxonomy of WG functions has the attraction of open-endedness and flexibility, in contrast with the HPSG analysis which assumes a fixed and universal list of dependency types defined by the order of elements in the various categories discussed previously (SPR, COMPS and ARG-ST). A universal list of categories requires an explanation: Why a universal list? Why this particular list? How does the list develop in a learner's mind? In contrast, a taxonomy can be learned entirely from experience, can vary across languages, and can accommodate any amount of minor variation. Of these three attractions, the easiest to illustrate briefly is the third. Take once again the English preposition *from*, as in (30).

(30) From London to Edinburgh is four hundred miles.

Here *from* seems to have two complements: *London* and *to Edinburgh*. Since they have different properties, they must be distinguished, but how? The easiest and arguably correct solution is to create a special dependency type just for the second complement of *from*. This is clearly unproblematic in the flexible WG approach, where any number of special dependency types can be added at the foot of the taxonomy, but much harder if every complement must fit into a universal list. (Note that the availability of a dustbin category such as "oblique" would not help, because the second complement of *from* has specific syntactic and semantic properties which need to be specified. If a category could be described as a dustbin, it is not a category because it does no work in the inheritance system.)

To summarise the discussion, therefore, HPSG and WG offer fundamentally different treatments of grammatical functions with two particularly salient differences. In the treatment of adjuncts, there are reasons for preferring the WG approach in which adjuncts and arguments are grouped together explicitly as dependents. But in distinguishing different types of complement, the HPSG lists seem to complement the taxonomy of WG, each approach offering different benefits. This is clearly an area needing further research.

6 HPSG without PS?

This chapter on HPSG and DS raises a fundamental question for HPSG: does it really need PS? Most introductory textbooks present PS as an obvious and established approach to syntax, but it is only obvious because these books ignore the DS alternative: the relative pros and cons of the two approaches are rarely assessed. Even if PS is in fact better than DS, this can't be described as "established" (in the words of one of my reviewers) until its superiority has been demonstrated. This hasn't yet happened. The historical sketch showed very clearly that nearly two thousand years of syntactic theory assumed DS, not PS, with one exception: the subject-predicate analysis of the proposition (later taken to be the sentence). Even when PS was invented by Bloomfield, it was combined with elements of DS, and Chomsky's PS, purified of all DS elements, only survived from 1957 to 1970.

A reviewer also argues that HPSG is vindicated by the many large-scale grammars that use it. These grammars are indeed impressive, but DS theories have also been implemented in the equally large-scale projects listed in Section 2. In any case, the question is not whether HPSG is a good theory, but rather whether it might be even better without its PS assumptions. The challenge for HPSG, then, is to explain why PS is a better basis than DS. The debate has hardly started, so its outcome is unpredictable; but suppose the debate favoured DS. Would that be the end of HPSG? Far from it. It could survive almost intact, with just two major changes.

The first would be in the treatment of grammatical functions. It would be easy to bring all dependents together in a list called DEPS (Bouma et al. 2001) with ADJUNCTS and COMPS as sub-lists, or even with a separate subcategory for each sub-type of dependent (Hellan 2017).

The other change would be the replacement of phrasal boxes by a single list of words. (31) gives a list for the example with which we started (with round and curly brackets for ordered and unordered sets, and a number of sub-tokens for each word):

(31) (*many*, *many+h students*, *students+a*, *enjoy*, *enjoy+o*, *enjoy+s*, *syntax*)

Each word in this list stands for a whole box of attributes which include syntactic dependency links to other words in the list. The internal structure of the boxes would otherwise look very much like standard HPSG, as in the schematic neo-HPSG structure in Figure 17. (To improve readability by minimizing crossing lines, attributes and their values are separated as usual by a colon, but may

appear in either order.)

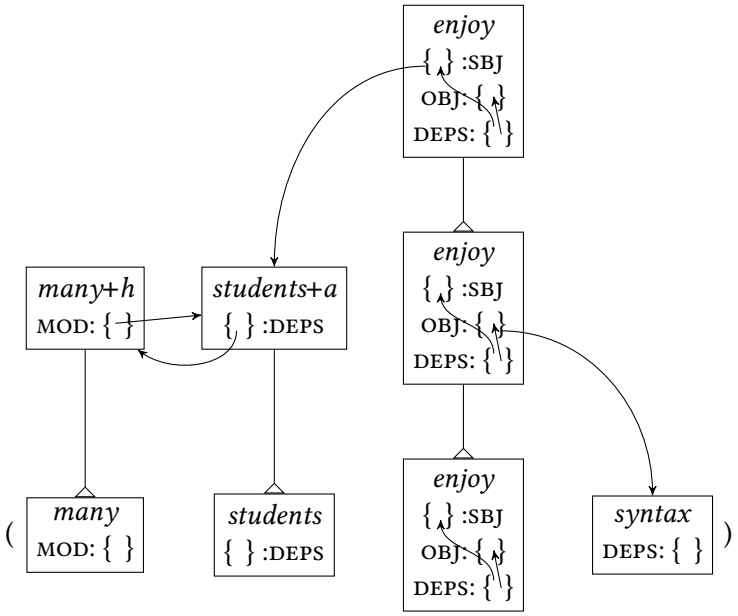

Figure 17: A neo-HPSG analysis

Figure 17 can be read as follows:

- The items at the bottom of the structure (*many*, *students*, *enjoy* and *syntax*) are basic types stored in the grammar, available for modification by the dependencies. These four words are the basis for the ordered set in (31), and shown here by the round brackets, with the ordering shown by the left-right dimension. This list replaces the ordered partonomy of HPSG.
- Higher items in the vertical taxonomy are tokens and sub-tokens, whose names show the dependency that defines them (*h* for "head", *a* for "adjunct", and so on). The taxonomy above *enjoy* shows that *enjoy+s* isa *enjoy+o* which isa *enjoy*, just as in an HPSG structure where each dependent creates a new representation of the head by satisfying and cancelling a valency need and passing the remaining needs up to the new representation.
- The taxonomy above *students* shows that *students+a* is a version of *students* that results from modification by *many*, while the parallel one above *many* shows that (following HPSG practice) *many+h* has the function of modifying *students*.

Roughly speaking, each boxed item in this diagram corresponds to an AVM in a standard HPSG analysis.

In short, modern HPSG could easily be transformed into a version of DS, with a separate AVM for each word. As in DS, the words in a sentence would be represented as an ordered list interrelated partly by the ordering and partly by the pairwise dependencies between them. This transformation is undeniably possible. Whether it is desirable remains to be established by a programme of research and debate which will leave the theory more robust and immune to challenge.

Acknowledgements

I would like to take this opportunity to thank Stefan Müller for his unflagging insistence on getting everything right.

References

Abeillé, Anne & Robert D. Borsley. 2008. Comparative correlatives and parameters. *Lingua* 118(8). 1139–1157. DOI: 10.1016/j.lingua.2008.02.001.

Abeillé, Anne & Rui Chaves. 2021. Coordination. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 697–748. Berlin: Language Science Press. DOI: ??.

Arnold, Doug & Robert D. Borsley. 2014. On the analysis of English exhaustive conditionals. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 27–47. Stanford, CA: CSLI Publications. http://csli-publications.stanford.edu/HPSG/2014/arnold-borsley.pdf (18 August, 2020).

Bender, Emily M. 2008. Radical non-configurationality without shuffle operators: An analysis of Wambaya. In Stefan Müller (ed.), *Proceedings of the 15th International Conference on Head-Driven Phrase Structure Grammar*, 6–24. Stanford, CA: CSLI Publications. http://csli-publications.stanford.edu/HPSG/2008/bender.pdf (19 November, 2020).

Berwick, Robert C., Angela D. Friederici, Noam Chomsky & Johan J. Bolhuis. 2013. Evolution, brain, and the nature of language. *Trends in Cognitive Sciences* 17(2). 89–98. DOI: 10.1016/j.tics.2012.12.002.

Billroth, Johann. 1832. *Lateinische Syntax: Für die obern Klassen Gelehrter Schulen*. Leipzig: Weidmann.

Blake, Barry. 1990. *Relational grammar*. London: Croom Helm.

Blevins, James & Ivan Sag. 2013. Phrase Structure Grammar. In Marcel den Dikken (ed.), *The Cambridge handbook of generative syntax*, 202–225. Cambridge, UK: Cambridge University Press. DOI: 10.1017/CBO9780511804571.010.

Borsley, Robert D., Maggie Tallerman & David Willis. 2007. *The syntax of Welsh* (Cambridge Syntax Guides 7). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486227.

Bouma, Gosse, Robert Malouf & Ivan A. Sag. 2001. Satisfying constraints on extraction and adjunction. *Natural Language & Linguistic Theory* 19(1). 1–65. DOI: 10.1023/A:1006473306778.

Bresnan, Joan. 1978. A realistic Transformational Grammar. In Morris Halle, Joan Bresnan & George A. Miller (eds.), *Linguistic theory and psychological reality*, 1–59. Cambridge, MA: MIT Press.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. 1st edn. Oxford: Blackwell Publishers Ltd.

Briscoe, Ted, Ann Copestake & Valeria de Paiva (eds.). 1993. *Inheritance, defaults, and the lexicon*. Cambridge, UK: Cambridge University Press.

Chen, Danqi & Christopher Manning. 2014. A fast and accurate dependency parser using neural networks. In Alessandro Moschitti, Bo Pang & Walter Daelemans (eds.), *Proceedings of EMNLP (Empirical Methods on Natural Language Processing) 2014*, 740–750. Stroudsburg, PA: The Association for Computational Linguistic.

Chomsky, Noam. 1957. *Syntactic structures* (Janua Linguarum / Series Minor 4). The Hague: Mouton.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick A. Jacobs & Peter S. Rosenbaum (eds.), *Readings in English Transformational Grammar*, chap. 12, 184–221. Waltham, MA/Toronto/London: Ginn & Company.

Crysmann, Berthold. 2008. An asymmetric theory of peripheral sharing in HPSG: Conjunction reduction and coordination of unlikes. In Gerhard Jäger, Paola Monachesi, Gerald Penn & Shuly Wintner (eds.), *Proceedings of FGVienna: The 8th conference on Formal Grammar*, 47–62. Stanford, CA: CSLI Publications.

Culicover, Peter W. & Ray Jackendoff. 1999. The view from the periphery: The English comparative correlative. *Linguistic Inquiry* 30(4). 543–571. DOI: 10.1162/002438999554200.

Dahl, Östen. 1980. Some arguments for higher nodes in syntax: A reply to Hudson's 'constituency and dependency'. *Linguistics* 18(5–6). 485–488. DOI: 10.1515/ling.1980.18.5-6.485.

Dik, Simon. 1989. *The theory of functional grammar: The structure of the clause*. Dordrecht: Foris.

Duran-Eppler, Eva. 2010. *Emigranto: The syntax of German-English code-switching*. Manfred Markus, Herbert Schendl & Sabine Coelsch-Foisner (eds.). Vol. 99 (Austrian Studies In English). Vienna: Wilhelm Braumüller Universitäts-Verlagsbuchhandlung.

Duran-Eppler, Eva, Adrian Luescher & Margaret Deuchar. 2017. Evaluating the predictions of three syntactic frameworks for mixed determiner–noun constructions. *Corpus Linguistics and Linguistic Theory* 13(1). 27–63. DOI: 10.1515/cllt-2015-0006.

Fillmore, Charles J. 1986. Varieties of conditional sentences. In *Eastern states conference on linguistics*, vol. 3, 163–182.

Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum & Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Cambridge, MA: Harvard University Press.

Gisborne, Nikolas. 2008. Dependencies are constructions. In Graeme Trousdale & Nikolas Gisborne (eds.), *Constructional approaches to English grammar*, 219–56. New York: Mouton.

Gisborne, Nikolas. 2010. *The event structure of perception verbs*. Oxford: Oxford University Press.

Gisborne, Nikolas. 2011. Constructions, word grammar, and grammaticalization. *Cognitive Linguistics* 22. 155–182.

Gisborne, Nikolas. 2019. Word grammar morphology. In Francesca Mansini & Jenny Audring (eds.), *Oxford handbook of morphological theory*, 327–45. New York: Oxford University Press. DOI: 10.1093/oxfordhb/9780199668984.001.0001.

Gisborne, Nikolas. 2020. *Ten lectures on event structure in a network theory of language*. To appear. Leiden: Brill.

Halliday, Michael A. K. 1961. Categories of the theory of grammar. *Word* 17. 241–282.

Halliday, Michael A. K. 1967. Notes on transitivity and theme in English. Part II. *Journal of Linguistics* 3(2). 199–244.

Harman, Gilbert. 1963. Generative grammars without transformation rules: A defence of phrase structure. *Language* 39. 597–616.

Hellan, Lars. 2017. A design for the analysis of bare nominalizations in Norwegian. In Maria Bloch-Trojnar & Anna Malicka-Kleparska (eds.), *Aspect and valency in nominals*, vol. 134 (Studies in Generative Grammar), 181–99. Berlin: Walter de Gruyter. DOI: 0.1515/9781501505430-008.

Hudson, Richard. 1982. Incomplete conjuncts. *Linguistic Inquiry* 13(3). 547–550.

Hudson, Richard. 1984. *Word grammar*. Oxford: Basil Blackwell.

Hudson, Richard. 1988. Coordination and grammatical relations. *Journal of Linguistics* 24(2). 303–342.

Hudson, Richard. 1990. *English Word Grammar*. Oxford: Basil Blackwell.
Hudson, Richard. 1995. Does English really have case? *Journal of Linguistics* 31. 375–392.
Hudson, Richard. 1998. *English grammar*. London: Routledge.
Hudson, Richard. 2007. *Language networks: The new Word Grammar*. Oxford: Oxford University Press.
Hudson, Richard. 2010. *An introduction to Word Grammar* (Cambridge Textbooks in Linguistics). Cambridge, UK: Cambridge University Press.
Hudson, Richard. 2016. Dependency Grammar. In Andrew Hippisley & Gregory Stump (eds.), *Cambridge handbook of morphological theory*, 657–682. Cambridge: Cambridge University Press.
Hudson, Richard. 2017. French pronouns in cognition. In Andrew Hippisley & Nikolas Gisborne (eds.), *Defaults in morphological theory*, 114–150. Oxford: Oxford University Press.
Hudson, Richard. 2018. Pied piping in cognition. *Journal of Linguistics* 54(1). 85–138. DOI: 10.1017/S0022226717000056.
Imrényi, András. 2013. Constituency or dependency? Notes on Sámuel Brassai's syntactic model of Hungarian. In Péter Szigetvári (ed.), *Papers presented to László Varga on his 70th birthday*, 167–182. Budapest: Tinta.
Imrényi, András & Zsuzsa Vladár. 2020. Sámuel Brassai in the history of Dependency Grammar. In András Imrényi & Nicolas Mazziotta (eds.), *The origins of Dependency Syntax. A historical survey from Panini to Tesnière*. To appear. Amsterdam: John Benjamins Publishing Co.
Jackendoff, Ray S. 2008. Construction after Construction and its theoretical challenges. *Language* 84(1). 8–28.
Jespersen, Otto. 1937. *Analytic syntax*. University of Chicago Press.
Kahane, Sylvain. 2020. How dependency syntax appeared in the French Encyclopedia: From Buffier (1907) to Beauzée (1765). In András Imrényi & Nicolas Mazziotta (eds.), *The origins of dependency syntax. A historical survey from Panini to Tesnière*. To appear. Amsterdam: John Benjamins Publishing Co.
Kathol, Andreas. 2000. *Linear syntax*. New York, NY: Oxford University Press.
Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8(1). 63–99.
Kern, Franz. 1884. *Grundriß der Deutschen Satzlehre*. Berlin: Nicolaische Verlags-Buchhandlung.
Kouloughli, Djamel-Eddine. 1999. Y a-t-il une syntaxe dans la tradition Arabe. *Histoire Épistémologie Langage* 21. 45–64.

Kübler, Sandra, Ryan McDonald & Joakim Nivre. 2009. *Dependency Parsing* (Synthesis Lectures on Human Language Technologies 2). San Rafael, U.S.A: Morgan & Claypool Publishers. 1–127.

Lambrecht, Knud. 1990. ''What, me worry?" — 'Mad Magazine sentences' revisited. In *Proceedings of the 16th Annual Meeting of the Berkeley Linguistics Society*, 215–228. Berkeley, CA: Berkley Linguistics Society.

Lascarides, Alex & Ann Copestake. 1999. Default representation in constraint-based frameworks. *Computational Linguistics* 25(1). 55–105.

Law, Vivien. 2003. *The history of Linguistics in Europe: From Plato to 1600*. Cambridge: Cambridge University Press.

Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago, IL: University of Chicago Press.

Luhtala, Anneli. 1994. Grammar, early mediaeval. In Ronald Asher (ed.), *Encyclopedia of language and linguistics*, 1460–1468. Oxford: Pergamon.

Marneffe, Marie-Catherine de, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre & Christopher Manning. 2014. Universal Stanford dependencies: A cross-linguistic typology. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)*, 4585–4592. Reykjavik, Iceland: European Language Resources Association (ELRA).

Matthews, Peter. 1993. *Grammatical theory in the United States from Bloomfield to Chomsky*. Cambridge: Cambridge University Press.

Mel'čuk, Igor A. 1988. *Dependency Syntax: Theory and practice* (SUNY Series in Linguistics). Albany, NY: SUNY Press.

Müller, Stefan. 1995. Scrambling in German – Extraction into the *Mittelfeld*. In Benjamin K. T'sou & Tom Bong Yeung Lai (eds.), *Proceedings of the Tenth Pacific Asia Conference on Language, Information and Computation*, 79–83. City University of Hong Kong.

Müller, Stefan. 1996. The Babel-System: An HPSG fragment for German, a parser, and a dialogue component. In *Proceedings of the Fourth International Conference on the Practical Application of Prolog*, 263–277. London.

Müller, Stefan. 2012. *On the copula, specificational constructions and type shifting*. Ms. Freie Universität Berlin.

Müller, Stefan. 2018. *Grammatical theory: From Transformational Grammar to constraint-based approaches*. 2nd edn. (Textbooks in Language Sciences 1). Berlin: Language Science Press. DOI: 10.5281/zenodo.1193241.

Müller, Stefan. 2019. Evaluating theories: Counting nodes and the question of constituency. *Language Under Discussion* 5(1). 52–67. DOI: 10.31885/lud.5.1.226.
Müller, Stefan. 2021. Constituent order. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 353–398. Berlin: Language Science Press. DOI: ??.
Nordlinger, Rachel. 1998. *A grammar of Wambaya, northern territory (Australia)*. Vol. C–140 (Pacific Linguistics). Canberra: Research School of Pacific & Asian Studies, The Australian National University.
Owens, Jonathan. 1988. *The foundations of grammar: An introduction to mediaeval Arabic grammatical theory*. Amsterdam: John Benjamins Publishing Co.
Percival, Keith. 1976. On the historical source of immediate constituent analysis. In James McCawley (ed.), *Notes from the linguistic underground*, 229–242. London: Academic Press.
Perlmutter, David M. & Paul M. Postal. 1983. Some proposed laws of basic clause structure. In David M. Perlmutter (ed.), *Studies in Relational Grammar*, vol. 1, 81–128. Chicago, IL: The University of Chicago Press.
Pike, Kenneth. 1954. *Language in relation to a unified theory of the structure of human behavior*. Glendale, CA: Summer Institute of Linguistics.
Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar* (Studies in Contemporary Linguistics). Chicago: The University of Chicago Press.
Reape, Mike. 1994. Domain union and word order variation in German. In John Nerbonne, Klaus Netter & Carl Pollard (eds.), *German in Head-Driven Phrase Structure Grammar* (CSLI Lecture Notes 46), 151–198. Stanford, CA: CSLI Publications.
Reed, Alonzo & Brainerd Kellogg. 1877. *Higher lessons in English. A work on English grammar and composition, in which the science of the language is made tributary to the art of expression. A course of practical lessons carefully graded, and adapted to every day use in the school-room*. New York: Clark & Maynard.
Reisberg, Daniel. 2007. *Cognition. Exploring the science of the mind*. New York City, NY: Norton.
Robins, Robert. 1967. *A short history of linguistics*. London: Longman.
Rosch, Eleanor. 1973. Natural categories. *Cognitive Psychology* 4(3). 328–350.
Rosta, Andrew. 2008. Antitransitivity and constructionality. In Graeme Trousdale & Nikolas Gisborne (eds.), *Constructional approaches to English grammar*, 187–218. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110199178.3.187.
Sériot, Patrick. 2004. L'affaire du petit drame: Filiation franco-russe ou communauté de pensée? (Tesnière et Dmitrievskij). *Slavica Occitania* 17. 93–118.

Sgall, Petr, Eva Hajičová & Jarmilá Panevova. 1986. *The meaning of the sentence in its semantic and pragmatic aspects*. Prague: Academia.
Siewierska, Anna. 1991. *Functional grammar*. London: Routledge.
Sugayama, Kensei. 2002. *Studies in Word Grammar*. Vol. 55 (Annals of Foreign Studies). Kobe: Research Institute of Foreign Studies, Kobe City University of Foreign Studies.
Sweet, Henry. 1891. *A new English grammar: Logical and historical*. Vol. 2. Oxford: Clarendon Press.
Tallerman, Maggie. 2009. Phrase structure vs. dependency: The analysis of Welsh syntactic soft mutation. *Journal of Linguistics* 45(1). 167–201.
Taylor, John. 1995. *Linguistic categorization: Prototypes in linguistic theory*. 2nd edn. Oxford: Clarendon.
Tesnière, Lucien. 1959. *Eléments de syntaxe structurale*. Paris: Librairie C. Klincksieck.
Tesnière, Lucien. 2015. *Elements of structural syntax*. Translated by Timothy Osborne and Sylvain Kahane. Amsterdam: John Benjamins Publishing Co.
Traugott, Elizabeth & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.
Zwicky, Arnold. 1985. The case against plain vanilla syntax. *Studies in the Linguistic Sciences* 15(2). 205–225.