

HPSG and Dependency Grammar

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, the bulk of this chapter is devoted to the history of syntactic theory and to the development of these two traditions.

The basis for PS analysis is the part-whole relation between smaller units (including words) and 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. Trees 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 AVMs, each box represents a unit which bears properties such as a classification, a meaning and relations to other items.

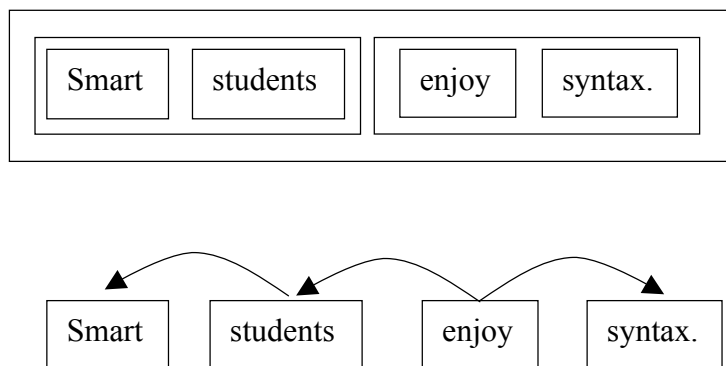


Figure 1: Phrase structure and dependency structure contrasted

Turning to Head-driven Phrase Structure Grammar, the term *Head-driven* points immediately to dependency: an asymmetrical relation

driven by a head word. On the other hand, *Phrase Structure* clearly locates the theory in the PS tradition. This chapter reviews the relations between HPSG and the very long DS tradition of grammatical analysis. The conclusion will be that it's easy to see HPSG as part of the dependency tradition; and it may not be a coincidence that one of the main power-bases of HPSG is Germany, where this tradition is also at its strongest (Müller 2018, 359).

But if HPSG recognises dependencies, does it still need PS? This is a difficult and surprisingly complicated question which, speaking personally, I have been struggling with most of my working life. My first encounter with syntactic theory was probably typical, with PS as the only approach mentioned. PS lies at the heart of the first theory I adopted: 'Systemic Grammar', now called 'Systemic Functional Grammar' (Halliday 1985; Hudson 1971). I studied and researched syntax for a decade before I even heard of DS, and at that point I concluded that maybe both DS and PS were needed (Hudson 1976); but a few years later I decided that DS could handle everything (Hudson 1984). DS was (to me) obviously right and needed; but if I had DS, why would I need PS as well? This is a line that I've pursued since then through the theory of Word Grammar .

Consequently, when HPSG arrived in 1994 (Pollard and Sag 1994), I celebrated the H but regretted the PS. Indeed, I even 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* could also be taken as 'Pollard and Sag'. Carl and Ivan took it well, and I think at least one of them may even have entertained the possibility that I might be right.

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. Here too I found welcome similarities, not least the use of default inheritance. I shall argue below that inheritance offers a novel solution to one of the outstanding challenges for the dependency tradition.

The next six sections set the historical scene. This is important because it's all too easy for students to get the impression 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 tradition in which the default is actually DS rather than PS. Later sections then address particular issues shared by HPSG and the dependency tradition.

2. The discovery of dependency relations

The Ancient World of Greece and Rome produced a large number of grammatical analyses of Greek and Latin, but the main focus of these grammars was the classification of words (in terms of word classes and inflectional categories such as gender, number, case, tense and so on). Syntax received much less attention, but since both languages had clear morphological case marking, it was hard not to notice the syntax of government: the fact that both prepositions and verbs ‘governed’ the case of an accompanying noun. Moreover, the same verb could govern the case of more than one such noun, so grammarians had to develop ways of distinguishing these nouns when discussing government. The obvious way to do this was in terms of the case selected, so dependent nouns were described as ‘the nominative’ or ‘the accusative’ rather than in terms of distinct relations – our ‘grammatical functions’ such as ‘subject’ and ‘object’. Other generalisations were described in terms of these case-defined nouns: agreement between the verb and its nominative, and transitivity selection between the verb and its ‘oblique cases’ (all but nominative) (Robins 1967, 29).

Both the selecting verb or preposition and the item carrying the case inflection were single words, so the 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, was first identified by the Arabic grammarian Sibawayh in the 8th century (Owens 1988; Kouloughli 1999). In the Arabic tradition which then developed, it was relatively easy for grammarians to recognise dependency because of two typological properties of Classical Arabic: three cases (which were extended to include three verbal moods) and fairly consistent head-initial word order. As a result, the dependency relation, including subjects as well as objects, was recognised not only as a bearer of case government but also as relevant to word order, and the metalanguage distinguished the relation of governor to governed from the particular case selected. In fact, the relation abstracted in this way was solid enough to allow a major debate between the two centres of

grammatical theory (Basra and Kufa) over whether mutual dependency was possible (Owens 1988, 52) – an issue to which I return in section 12. But even though grammatical theory recognised dependency as a relation, it only recognised one kind of dependency: that between a governor and a governed noun or verb. This dependency included subjects as well as objects and other complements, but it excluded adjuncts; so no relation was recognised between a noun and its modifying adjectives.

3. The invention of the subject-predicate analysis

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, it was obvious that 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 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 persists even today in some school grammars, and which 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 and verb were both equally essential – an analysis which even logicians have now abandoned in favour of a Fregian dependency analysis, but which still persists in some PS analyses and in some school grammars. Moreover the facts of subject-verb agreement, in which the subject determines the inflection of the verb, suggested that the noun was at least as important as the verb; logic treated them as equals, but the grammatical tradition includes a

sprinkling of analyses in which the noun dominates the verb. These analyses range from the modistic grammarians of the twelfth century (Robins 1967, 83), through a number of individual grammarians in the nineteenth century such as Solomon Barrett (Barrett 1845), 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'.

In short, the special status of the subject-verb relation seemed obvious to many, in contrast with all the other syntactic relations where DS was increasingly obvious. The first grammarians to recognise a consistent dependency-based analysis for the rest of the sentence were the French *encyclopédistes* of the eighteenth century (Kahane forthcoming), and by the nineteenth century much of Europe accepted a theory of sentence structure which by modern standards is muddled and complicated. In this theory, a sentence necessarily had a subject noun and a predicate; but (to increase the complication) the predicate could also be supported by a 'copula', so in a sentence like *Mary was tired*, the predicate would be *tired*, and *was* would be a third element in the top-level analysis. Each of these units was a single word, not a phrase, and the subject and predicate could be expanded by dependents. Function words such as prepositions had no proper position, being treated typically as though they were case inflections.

We might wonder why this hybrid analysis proved so popular. The reason is a matter of conjecture, but two explanations seem plausible: the perceived scope of syntax, and the available technology. Syntactic structure held words together, so the focus was on the single word. In teaching syntax, the main activity was 'parsing', a tradition dating from Priscian himself, and formalised in the ninth century (Luhtala 1994). (It's only recently that the term *parsing* has been applied to a sentence-level syntactic analysis.) 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 did parsing at school). In HPSG terms, the school child was asked to provide the values for the relevant attributes of a selected word, including its grammatical function (which would explain its case). In the early centuries the language was Latin, but more recently it was the vernacular (such as English). An English grammar that was particularly popular in Britain (Nesfield 1898, 107) lists the attributes of a noun:

- its kind (Proper, Common, Collective, Material or Abstract)

- its gender (Masculine, Feminine or Neuter)
- its number (Singular or Plural)
- its case (Nominative, Possessive or Objective)
- why in such a case (Nominative to verb and fourteen other possibilities).

None of these questions could be answered on the basis of rote learning, and given the paucity of formal markers, many required a sophisticated understanding of the sentence structure; but the focus was on the single word rather than the sentence.

In short, until recently grammar was focused on single words, and this was one of the reasons why the contradiction between the subject-predicate analysis and general dependency analysis could pass unnoticed. The other reason was the limited technology for discussing grammatical structure in print. Medieval scholarship was based on prose descriptions, and this continued to be the case after printing was invented. The early technology of printing made it difficult even to draw a straight vertical line, and it wasn't until the early nineteenth century that the idea of drawing a diagram to show the anatomy of a sentence occurred to anyone. For a structure as complicated as sentence structure, a diagrammatic notation is as essential as maps are for geography; and without diagrams, it is all too easy to overlook inconsistencies.

What seems to have been the very first syntactic diagram was published in 1832 in a Latin grammar for school children (Billroth 1832) – an unexpected place from a modern perspective, but quite natural in the days when pedagogy was the main application of grammar, and when Latin still dominated the school curriculum. The book was widely used, so although this diagram was the only one in the book, it may have had a great deal of influence (not least on Gottlob Frege (Hodges 2005), whose father had also written a school grammar of Latin (Frege 1850)). The diagram is reproduced in Figure 2. It shows the structure of the Latin sentence (1).

- (1) Miltiades dux Atheniensium reddidit libertatem paene
Miltiades leader of-Athenians restored freedom almost

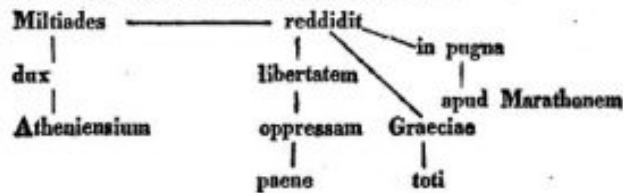
oppressam toti Graeciae in pugna apud Marathonem.
crushed to-all Greece in battle at Marathon.

(Latin)

'Miltiades the leader of the Athenians restored the freedom
which had almost been crushed to all Greece in the battle at

Marathon.'

§. 131. Auf diese Weise wird z. B. folgender Satz: *Miltiades, dux Atheniensium, toti Graeciae libertatem paene oppressam in pugna apud Marathonem reddidit*, dessen Körper eigentlich nur aus dem Subj.: *Miltiades* und dem Prädicat (mit der Copula) *reddidit* besteht, bekleidet:



better
quality?

Figure 2: A syntactic diagram from Billroth (1832)

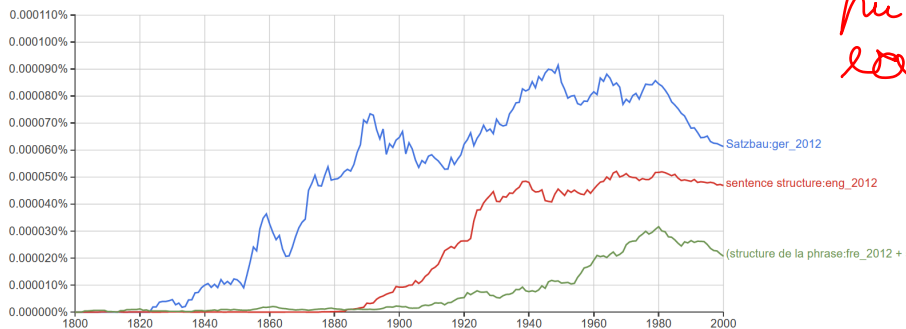
As Billroth says, the ‘body’ of the sentence actually consists of just two words: the subject *Miltiades* and the predicate (‘with the copula’) *reddidit*. The notation distinguishes these two words from the rest of the sentence by means of a horizontal line which contrasts with the vertical or diagonal lines showing dependents. The inconsistency in the analysis is now impossible to ignore, and is indeed built into the notation; and this isn’t the only inconsistency revealed in this way. Another is in the treatment of prepositions (*in*, *apud*), which are simply attached to their object noun as though they were inflections. The modern view that function words are often heads rather than dependents is one of the achievements of the late twentieth century. Without diagrams it would have been much easier to ‘fudge’ the analysis by ignoring these inconsistencies, and in a sense the story of syntactic theory since the early nineteenth century has been our attempts to resolve inconsistencies that are highlighted by our notations.

What was
‘the inconsistency’?

4. The discovery of sentence structure

The introduction of sentence diagrams is closely linked to another development in syntactic theory: the idea that a sentence has a global structure which can be analysed in parallel with the local structures of individual words – in other words, the discovery of ‘sentence structure’. This idea seems to be a surprisingly recent innovation. It was probably proposed first in eighteenth-century France (Seguin 1993), and then, apparently independently, in Germany (Becker 1831). The historical story can be read very easily off the Google n-

grams in Figure 3, showing the relative frequency of the French *structure de la phrase* or *strucure des phrases* (bearing in mind that the French word *phrase* means ‘sentence’), the German *Satzbau* and the English *sentence structure*.



published
earlier?
2012?
may be
update

Figure 3: 'Sentence structure' since 1800 in French, German and English

On closer investigation, this diagram shows that the notion of ‘sentence structure’ was already present in French books in 1800, but appeared in German books only in the 1820s and in English books not until the 1870s. There is no evidence for French influence on German writers, but English grammars built explicitly on the German ideas (as in the title of a book published in London: *The analysis of sentences explained and systematised, after Beckers' German grammar* (Morell 1855)).

provided
the google
database
⇒ balanced

The new focus on sentence structure introduced the possibility of a new approach to syntax in the classroom, called ‘analysis’ – a top-down approach which always started with the old subject-predicate distinction, and then identified the various chains of dependents. By the end of the nineteenth century, analysis had joined parsing as the two pillars of grammar teaching, and is best known from the American diagram system patented in 1877 by Reed and Kellogg (Reed and Kellogg 1877). Reed-Kellogg diagrams were enormously popular in the USA, and are still taught (under the simple name ‘diagramming’) in some American schools; indeed, there is a website (Sentence Diagrammer, by laiway) which generates them, giving diagrams such as the one in Figure 4. The significant feature of this diagram is, once again, the special treatment given to the relation between the subject and predicate (with the copula sitting uncomfortably between the two).

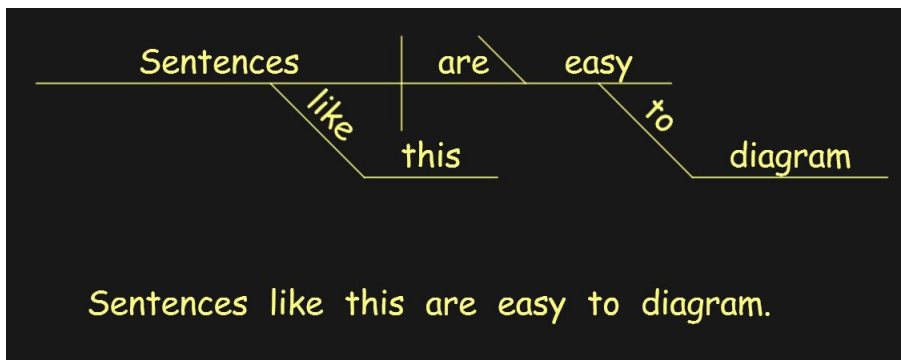


Figure 4: Reed and Kellogg diagram by Sentence Diagrammer

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.

The 1870s was an extraordinarily productive period in the history of syntactic theory, and we return to it below; but syntactic theory seems to be an area of intellectual life where the wheel is rediscovered over and over again, and where scholars cling to the familiar and ignore work in unfamiliar places or frameworks. This was as true in the nineteenth century as in the twenty-first, and does the field no credit. However, the silver lining of this cloud is that this history can treat different threads as independent, so we lose little by putting the parallel developments of the 1870s to one side while pursuing this strand of historical development.

I do not
understand
what is
meant
here
Rephrase?

5. The invention of phrase structure

In approaching the birth of Phrase Structure as we know it, our first encounter is with Leonard Bloomfield, father of Immediate Constituent Analysis. Given the history of diagramming in American schools, it seems almost certain that he learned about sentence analysis (as described above) 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 diagramable or not, had been applied in those grammars. (Percival 1976, 18)

However, he seems to have decided to ignore what he had (presumably) learned at school, just as he ignored the new approaches being applied in German schools when he visited Germany in 1913-14. Instead, he decided it was important to start from scratch, and especially so in approaching the exotic languages that he was studying in the USA. But we shall see that his school grammar may have influenced him in ways that he was unaware of.

For him, the main source of ideas about syntax was the German psychologist Wilhelm Wundt, who had evolved an alternative homogeneous theory of syntax, but this time one based entirely on the subject-predicate link (Percival 1976). The diagram in Figure 5 is quoted by Percival from Wundt's 1900 *Völkerpsychologie* of 1900; it represents Wundt's analysis of sentence (2).

- (2) Ein edlich denkender Mensch verschmäht die Täuschung.
 a sincerely thinking person scorns the deception

translation

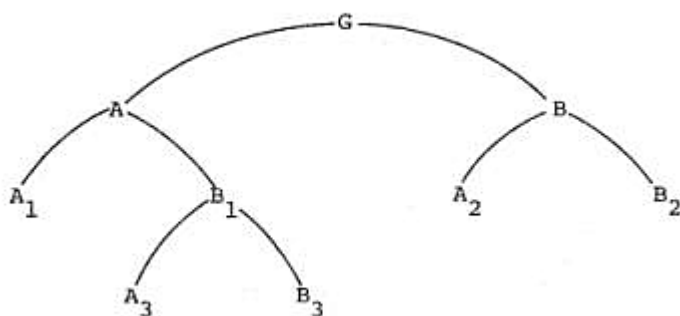


Figure 5: Wundt's diagram for Ein edlich denkender Mensch verschmäht die Täuschung

translation

In this diagram, G stands for *Gesamtvorstellung*, the 'total conceptualisation' of the sentence, consisting of a subject (A) and a predicate (B); and each of these elements in turn consists of another subject-predicate pair. For instance, A₁ is 'a person' and B₁ is 'thinks

sincerely', which in turn consists of 'thought' and 'is sincere'. For all its reliance on logic rather than grammar, the diagram is a clear precursor to neo-Bloomfieldian trees: it recognises a single consistent relationship which applies recursively. Moreover, this relationship is the part-whole relationship in which each level of structure gives an exhaustive analysis of the whole. This, then, is the start of the PS tradition: an analysis based purely on meaning and developed by a psychologist, not a grammarian. Wundt's theories of cognitive structure have long since lost credibility, so maybe it's time ~~for~~ the PS tradition to reconsider its roots.

for

Returning to the influence of Bloomfield's schooldays, it is easy to see the hybrid structures of Reed-Kellogg diagrams in his classification of structures as endocentric (based on dependency) or exocentric. 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. On the other hand, ICA paired these relations with the part-whole relation, which was homogeneous, so ICA combined a homogeneous PS analysis with an inconsistent DS analysis.

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 analyses 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 and Sag 2013, 202–3, my italics)

On the positive side, this quotation recognises DS as an alternative, but it implies, wrongly, that ICA discarded DS altogether.

Although Bloomfield avoided diagrams, his followers used them and taught students to use them, **though it was some time before modern tree diagrams were evolved (about a century after they were first invented in Europe)**. Two examples suffice. Figure 6 is from the first publication to use diagrams extensively in an ICA grammatical study

(Nida 1943, 18), and shows the continuing contrast between the exocentric subject-predicate structure (marked by 'x') and the endocentric structures whose head is indicated by the arrow-head. Figure 7 is from a widely used textbook (Hockett 1958, 188) which uses boxes rather than trees, but once again uses arrows to indicate dependency relations.

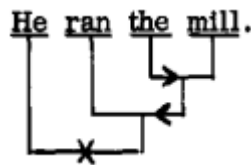


Figure 6: A syntactic tree by Nida

full reference here

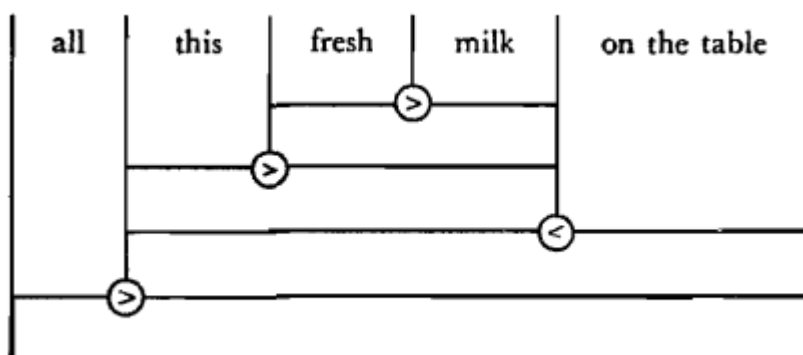


Figure 7: A box diagram for syntax

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 not for the traditional grammatical functions such as 'subject' and 'object'. Reacting against this gap, and presumably remembering their schoolroom training, some linguists developed syntactic theories which did have a place for grammatical functions, though not for dependency as such. The most famous of these theories are

- Tagmemics (Pike 1954)
- Functional Grammar (Dik 1989; Siewierska 1991)
- Systemic Functional Grammar (Halliday 1961, 1967)

lexical Functional frames

- Relational Grammar (Perlmutter and Postal 1983; Blake 1990)

The next player in the story is Noam Chomsky. As we all know, he defined phrase structure (PS) and presented it 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). Soon after, Paul Postal espoused this doubtful claim and developed it (Postal 1964), though he subsequently abandoned phrase structure in favour of grammatical relations in Relational Grammar.

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, like Bloomfield before him, seems to have forgotten what he probably learned at school about sentence analysis and diagramming; and (like many other linguists) he was unaware of the DS tradition. As a ^{which year?} graduate student he did study **Categorial Grammar, which he sees as an example of DS (pc)**, but ~~however important it is in the history of HPSG, Categorial Grammar is at best an untypical example of DS. Consequently, he dismissed both dependencies and grammatical functions as irrelevant clutter, leaving nothing but part-whole relations, sequential order and category-labels. Rather remarkably, the distinctive characteristic of his phrase structure was the (psychologically implausible) claim that ‘sideways’ relations such as dependency are impossible in syntax.~~ ^{But aren't these heads and dependents in X-bar syntax?}

However, it soon turned out that he had thrown out the relational baby with the bath water, and the dependency-free theory only lasted thirteen years. In 1970 he laid the grounds for X-bar theory (Chomsky 1970), which brought back dependency relations, combined with implicit grammatical functions. Explicit grammatical functions returned not only in Relational Grammar, but also in Lexical Functional Grammar (Bresnan 2001), though in this theory dependency-free phrase structure survives as well. ^{Bresnan & Kaplan, 1989 cite 1987, 1984 earliest reference}

The next important step in syntactic theory was the arrival of HPSG, which put dependencies and grammatical functions at the centre of syntactic structure (Pollard and Sag 1994). One of the attractions of HPSG for a dependency grammarian is the idea that syntactic structure is built out of head-dependent pairs. This idea was inherited from Categorial Grammar, and it is interesting to note that Categorial

Grammar can probably be traced back to Frege (Buszkowski 1989), who (as mentioned earlier) may have learned head-centred dependency analysis at school. I explore this idea and others in later sections.

6. The discovery of dependency structure



Having brought the history of syntax up to the present in the PS world, we now return to the 1870s in Europe to pursue the history of the DS world. As we have seen, both the rise of sentence-based syntactic analysis (as opposed to word-based parsing) and the development of efficient diagramming methods had offered a challenge to syntactic theory: how to reconcile the inconsistencies in the hybrid model which combines the subject-predicate relation with dependency relations. Most of syntactic theory perpetuated the inconsistency, and continues to do so wherever Reed-Kellogg diagrams are taught, but major thinkers were attracted in the nineteenth century to syntactic theory, and in particular to the search for a unified theory.

Rather remarkably, three different grammarians seem at roughly the same time to have independently reached the same conclusion: hybrid structures can be replaced by a homogeneous structure if we take the finite verb as the root of the whole sentence. This idea seems to have been proposed in print in 1873 by the Hungarian Sámuel Brassai (Imrényi 2013); in 1877 by the Russian Aleksej Dmitrievsky (Sériot 2004); and in 1884 by the German Alexander Kern (Kern 1884).

Dmitrievskij didn't use diagrams, but the other two did, so Brassai's *Figure 8* may be the first ever diagram which can be described as a 'dependency diagram'. The bracketed Hungarian terms indicate grammatical functions; for example, *uralk(odó) ige* means 'governing verb', but once again the sentence analysed is in Latin:

- (3) Uxor amans flentem flens acrius ipsa tenebat,
 wife loving crying crying more.bitterly herself was.hugging

imbres per indignas usque cadente genas.
 shower on unbecoming continuously falling cheeks

'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]'.
 (punctuation before ')

state
 language
 in brackets

(Latin)

no
 italics

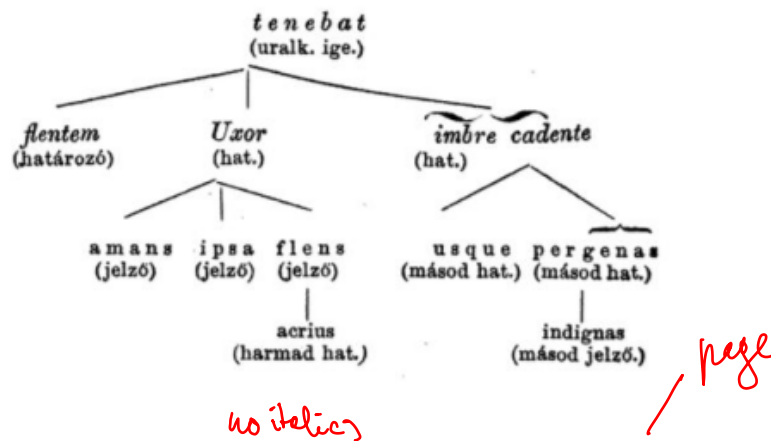


Figure 8: A verb-rooted tree from Brassai (1873)

However, even though we've now lost the horizontal subject-predicate link, the rather awkward horizontal brace over *imbre cadente* (two ablative cases in agreement, meaning 'with a shower falling') reminds us that the analysis still hasn't achieved the degree of homogeneity that we expect in modern analyses.

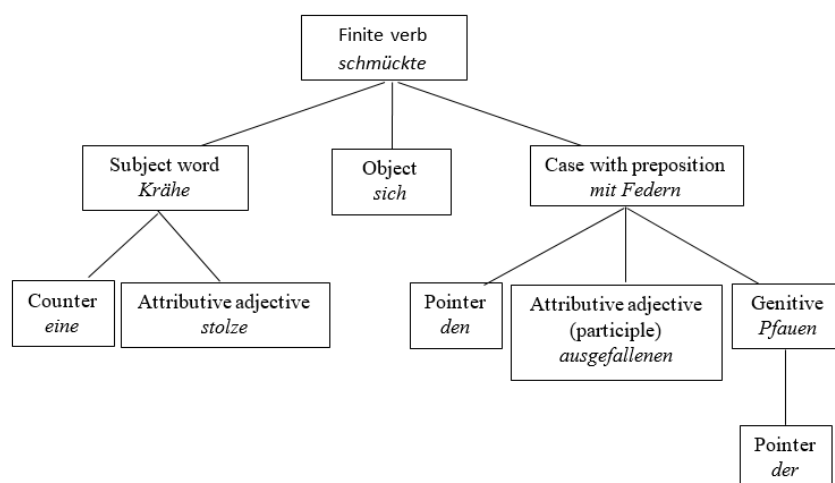
Just eleven years later, Kern published the diagram in Figure 9, which again shows the finite verb as the root of the sentence, giving (at least in principle) a consistent dependency basis for the whole analysis. The diagram has been redrawn for clarity, and the grammatical terminology translated into English, but the example is German:

- (4) Eine stolze Krähe schmückte sich mit den
a proud crow decorated himself with the

ausgefallenen Federn der Pfauen.
fallen-out feathers of the peacocks

'A proud crow decorated himself with the dropped feathers of the peacocks.'

ü ??



Cool

Figure 9: A verb-rooted tree from Kern (1884)

Figure 9, like Brassai's earlier one, shows the finite verb as the sentence root, but also shows the same ambivalence about 'awkward' structures, in this case the preposition *mit*, 'with', treated as a single element with the noun in spite of being separated from it in the phrase *mit den ausgefallenen Federn der Pfauen*.

By the end of the nineteenth century, therefore, European grammarians had already invented homogeneous syntax – the idea that a syntactic theory could postulate a single type of relation – and in all these theories, the single consistent relationship was dependency (with subtypes such as subject and object) and the hallmark was the dominating finite verb with the subject as one of its dependents alongside the object and other complements. Moreover, the basic unit of syntax was the word, not the sentence:

... 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)

In short, syntax at least aspired to pure dependency structure.

7. Variations on the dependency theme

The rest of the DS story is better known. 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’cuk 1988), who combined DS with multiple analytical levels, and the Czech linguists Petr Sgall and Eva Hajičová (Sgall, Hajicová, and Panevova 1986) who included information structure. It is noticeable that these theories have geographical roots in Eastern Europe, and that Tesnière was a slavist. Central Europe also figures in the DS story, represented in the nineteenth century by Brassai in Hungary and Kern in Germany, and DS is still relatively popular and well known there compared with the USA and other anglophone countries.

f

However, dependency structure has had a much wider geographical appeal in natural-language processing (Kübler, McDonald, and Nivre 2009). For example, at the time of writing (July 2018):

- According to Osborne and Kahane, translators of Tesnière’s book, “dependency parsing has really boomed, becoming more popular than constituency-based parsing”(Tesnière 1959, lxviii).
- the Wikipedia entry for ‘Treebank’ classified 50 of its 101 treebanks as using dependency structure.
- The ‘Universal dependencies’ website lists more than 100 dependency-based treebanks for 60 languages.
- Google’s n-gram facility uses dependency structure (and allows searches based on dependencies).
- The Stanford Parser (Chen and Manning 2014; de Marneffe et al. 2014) uses dependency structure.

2 ? n-grams
are
n-grams
not dependen
cies

The attraction of DS is that all the units of analysis (words) are given in the raw data, so at least in principle the overall analysis can immediately be broken down into a much simpler analysis for each word. This is as true for a computer scientist 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.

On the other hand, the theoretical and descriptive issues that divide PS grammarians also apply to those who use DS. On the descriptive side, all the 'awkward' constructions mentioned above receive different treatments from different grammarians. For example, when a preposition combines with a following noun, which is head? In the Stanford Parser, the preposition depends on the noun, but in most other DS analyses, as in most PS analyses, the preposition is head. Similar divisions apply to auxiliary verbs in relation to the following verb: dependents in some analyses, heads in others.

Equally profound disagreements affect the theoretical underpinnings of DS. 'Plain-vanilla' (Zwicky 1985) versions of DS and PS are very similar and are weakly equivalent (Hays 1964; Robinson 1970), but as with PS, such theories need to be supplemented, giving rise to theories in which structures are much richer. One of many such areas of disagreement concerns word order: should a theory require 'projectivity'? A DS analysis is projective if a line would project from each word token to its node in the analysis without crossing any dependency lines – in other words, if the phrases are all continuous. The challenge for projectivity is of course the existence of discontinuous phrases, as in raising or extraction examples such as (5) and (6).

(5) It is raining. [*it* depends on *raining* as well as on *is*.]

(6) What are you planning to do? [*what* depends on *do* as well as on *are*]

One reaction to such examples is to abandon projectivity (Nivre 2006), but others preserve projectivity by adapting the dependency structure (as explained below).

Another theoretical issue concerns the possibility of mutual dependency. This is the issue that set Basra against Kufa in the ninth century (see section 1.2), but it still divides theoreticians (though very few even consider the possibility of mutual dependency). Once again, we shall return to this issue below.

Since this book is about HPSG, I shall now focus on the DS theory which is probably nearest to HPSG. This will allow a few suggestions for directions in future HPSG research.

8. Word Grammar and structure sharing

The DS theory in question is my own theory Word Grammar (WG),

the ?

does it?
why?

Reap-style, see discussion
in 5th book.

(Hudson 1984, 1990, 2007, 2010; Eppler 2004; Gisborne 1996). The most distinctive characteristic of WG, compared with other DS theories, is probably the richness of its structures. Because the theory aims at cognitive plausibility, and because complex networks are so pervasive in general cognition, the theory allows the structure of a sentence to be a network, without formal limits on its possible complexity.

One consequence of the network approach is that a word is allowed to depend on more than one other word. The resulting structures are exactly parallel to those in HPSG that involve structure-sharing, described as 'the most important means of expression in HPSG' (Müller 2018, 267). For example, consider example (7).

(7) What did you see?

The relevant challenge in this example is the extraction of *what*, which is the object of *see* so in some sense (to be discussed below) its expected position is after this verb. Figure 10 offers two analyses both of which involve structure-sharing, one in the spirit of HPSG and the other in a simplified version of WG. In the HPSG analysis, the structure sharing is shown by coindexing *what* to a trace in its expected position, while the WG analysis shows it as a double dependency, with *what* depending both on *did* and on *see*. (The function labels can be read as 'extractee', 'subject' and 'extractee and object'.) As it happens, the example also involves raising, so *you* is shared as subject by both *see* and *did*; but I haven't tried to guess how this would be handled in HPSG.

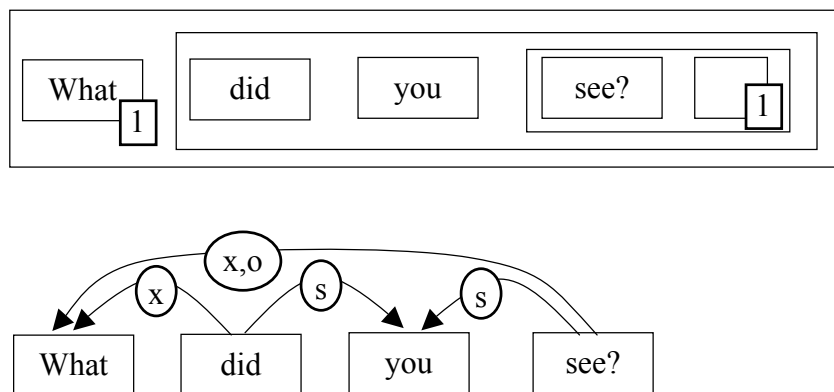


Figure 10: Extraction in HPSG and WG

Didn't
I also say
this in the
HPSG/Dy
comparison.

o /b

you could
read about
it.
It is struc-
ture
sharing as
well.

However, the structure-sharing analysis raises a question: if *what* could stand after *see*, why doesn't it stand there? Why do both raising and extraction always raise, rather than lowering? HPSG postpones this question by introducing a trace in the lower position, but for WG the answer invokes another important feature of the theory: default inheritance (which, of course, also plays an important part in HPSG). Somehow the analysis must show that the actual position of *what* overrides its default position, as object, after *see*. The WG solution recognises two distinct word-tokens for *what*, one functioning and positioned as extractee and the other as object. Let's call these two tokens *what-x* (the extractee *what*) and *what-o* (the object *what*, the equivalent of the HPSG trace). The relation between them is the 'isa' relation which carries default inheritance and which, in WG notation, is indicated by a small triangle whose base rests on the super-category. This analysis is shown in Figure 11, where *what-x* isa *what-o*. By the logic of default inheritance, the properties of the sub-category override those of the super-category, so the position of *what* is explained. This configuration is itself part of the grammar, and indeed it seems to be a default for grammars across languages, but default inheritance allows exceptions, and we do indeed seem to find exceptions such as German Partial VP Fronting (Müller 2018, 198; Hudson 2007, 143–44).

source
version
of

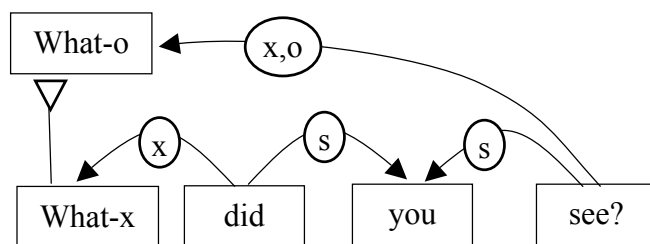


Figure 11: Extraction with default inheritance

This discussion of extraction has illustrated a significant similarity between HPSG and WG, but it has also introduced an important claim of the current version of WG which will offer a solution to a serious problem that dogs every DS theory. The claim is that a single word token may be factored into a number of simultaneous but distinct cognitive entities, as I did for *what*, where *what-x* and *what-o* have different attributes. The problem which this claim solves is discussed in the next section.

9. Semantic phrasing

One of the most obvious features of typical DS analyses is that every word has just one node in the analysis. One of the consequences is that the structure is completely flat, so if a word has more than one dependent, they are all attached directly to the same node. (The same is of course true of some PS analyses, such as those which recognise a flat structure for clauses where the subject and object are both immediate constituents.) This flat syntactic structure raises serious problems for the semantics, where ‘phrasing’ is required. 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 doesn’t 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 creates a new structure, semantic as well as syntactic, as it is combined with the growing phrase; and the same is true of Categorical Grammar, which provided the model for this principle. But for DS theories, this is not generally possible. Indeed, it is this characteristic of Categorical Grammar that makes it such a poor example of a DS theory.

Fortunately, however, these facts can be reconciled with the DS rejection of phrasal nodes (Hudson 2017). In the case of *typical French house*, let’s assume that an extra node is needed for *French house*, meaning ‘French house’. The only question is how this node is related to the node for *house*. In PS-based theories such as HPSG, there is a part-whole relation, but there is an alternative: the ‘isa’ relation introduced in the last section. In this analysis, *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* after modification by *French*, meaning ‘French house’.
- *house+t*: the word *house-F* after modification by *typical*, meaning ‘typical example of a French house’

The number of nodes in the 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. To show the significance of this difference, consider

the difference between the taxonomy 'novel' isa 'book' and the partonomy 'page' is-part-of 'book': a novel is not 'part of' a book, nor is a page a kind of book. The WG analysis of *typical French house* is shown in Figure 12.

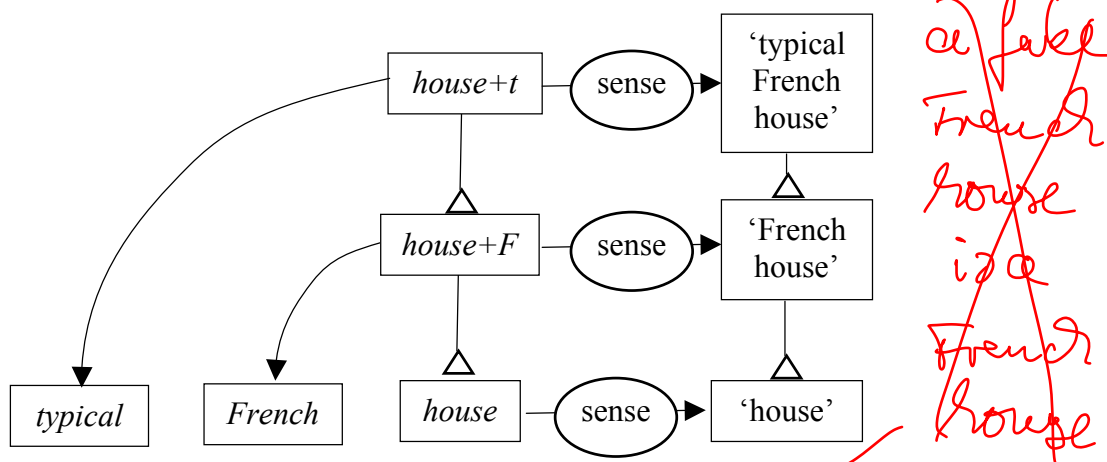


Figure 12: typical French house in WG

This analysis treats the word *house+F* as a particular kind of *house*, in just the same way that it treats 'French house' as a particular kind of 'house'. In both cases, the particular case inherits all the properties of the general except any that are overridden; so a small house is a typical house except for its size, and a fake house inherits at least some visual properties of a typical house, but not the main ones. **Since default inheritance is available in HPSG**, the same analysis could be introduced there without theoretical cost, as explained below.

not sure
it is under
the assumption
of Kiny & Riddle

10. Grammatical functions

One of the achievements of our long grammatical tradition is the development of a classification of dependents into what we now call 'grammatical functions' – 'subject', 'object' and so on. These are abstract categories which are based, like all other categories, on a multiplicity of converging properties; for example, a typical subject in English may be distinguished by its so-called case morphology, its position, its agreement with the verb, its availability for raising, its semantic role, and so on and on (Keenan 1976). But precisely because they are ordinary categories, they raise important questions in syntactic theory. The following list includes the WG answers:

- How do they relate to word-classes? General cognition

distinguishes two different kinds of category: entities and relations. For instance, in the social world the entities are people, and relations are all the social relations that we recognise, from 'parent' to 'classmate'. In the world of syntax, the entities are words, and the relations include dependencies and word-order, with grammatical functions as subtypes of 'dependent'.

- How many are there, and do they form a hierarchy? They constitute an open-ended and large taxonomy of categories in which more specific categories inherit from the more general but also have exceptional characteristics.
- Are they universal and innate, or parochial and learned? All categories are learned from experience (rather than innate), and the same is true of syntactic dependencies; and because of this, we can expect functional pressures to produce different subtypes of dependent in different languages.
- How should they be represented analytically? Like other categories, they are represented by labelled nodes in the overall network.

HPSG answers to these questions would probably be rather different, but the differences could be confronted with evidence.

For example, are grammatical functions categories (WG) or are they atomic values (HPSG)? This reduces to the question of whether they form a hierarchy, like entity categories; if they do, then they must themselves be categories. And there is plenty of evidence for a taxonomy of grammatical functions, with 'dependent' at the top of the hierarchy, running through general types such as 'valent' (subject or complement) and 'adjunct', to very specific categories like 'the second dependent of *from* which is realised as *to*, as in *from London to Birmingham*'.

} This is the HPSG handbook, so there should be answers.

? HPSG does not make use of grammatical functions.

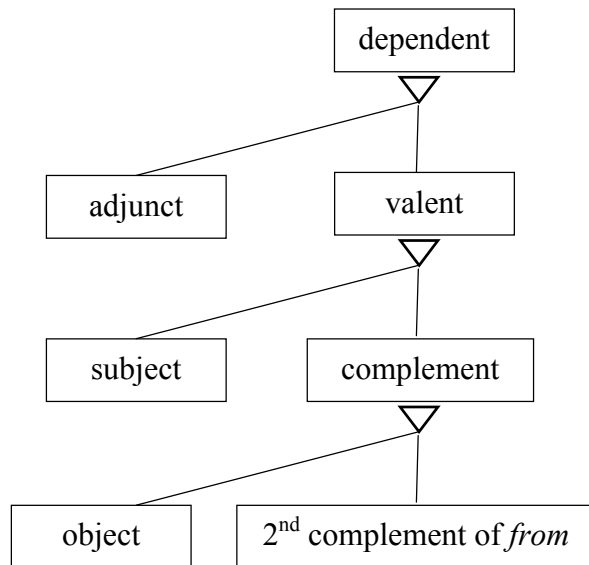


Figure 13: A taxonomy of grammatical functions

It is true that HPSG also offers a hierarchical analysis of grammatical functions, but the two hierarchies are very different. In contrast with the WG taxonomy, HPSG uses an ordered list in the COMPS list which ranks the valents (which may or may not include the subject) (Müller 2018, 266–67). The grammatical functions are only distinguished by their position in the list, so the functions themselves are certainly not explicit analytical categories which might themselves have properties such as ‘isa’ relations.

This hierarchy raises another issue which divides HPSG from WG: the general category ‘dependent’, which is central to WG but (apparently) not recognised in HPSG, where dependents are divided between complements (i.e. WG valents, including subjects) and adjuncts (Müller 2018, 281). The general category plays an important part in generalisations about word order such as those that distinguish head-final from head-initial languages, according to whether dependents typically precede or follow their heads. In WG such generalisations are easy – and indeed it is easy to see that languages like English constitute a third category of ‘head-medial languages’, where virtually every word-type allows dependents both before and after the head.

11. Word order

However the theme of this chapter is the similarity between WG and

HPSG, and an important similarity is in the treatment of word order, where both theories separate dominance (i.e. daughterhood in HPSG and dependency in WG) from linear precedence. In HPSG, the order of two elements is determined on the basis of their dominance properties, while WG predicts word order on the basis of dependency relations.

The WG mechanism for word order gives a great deal of flexibility as it uses a separate pair of relations: 'position' and 'landmark', each of which points to a node in the overall network (Hudson 2017). By default, a word's landmark is the word on which it depends, so this is normally the word from which it takes its position; this is what holds phrases together, because dependents keep as close to their landmarks as possible and a general principle bans intersecting landmark relations. The word's position is (of course) different from that of its landmark, because no two words can have the same position, so it is defined as either preceding or following the landmark – the WG equivalent of the LP operator '<' meaning 'precedes'. In any language, a word's landmark is determined by its dependency relations, but in a language such as English with relatively 'fixed' word order, the dependency also determines whether the word stands before or after its landmark.

Wangiri?

This mechanism brings word order within the scope of default inheritance, since a word's position is one of its properties which may be inherited or overridden. This can be seen easily if we return to the analysis of *What did you see?* by adding the new relations 'landmark' (lm) and 'position' (psn) to give Figure 14. This shows that both *what+o* and *what+x* have a landmark and a position, which in each case is inherited from the grammar; and that in each case the positions are related by '<', which shows which comes first. But since *what+x* is a *what+o*, its landmark and position override the default, thereby explaining the effects of extraction.

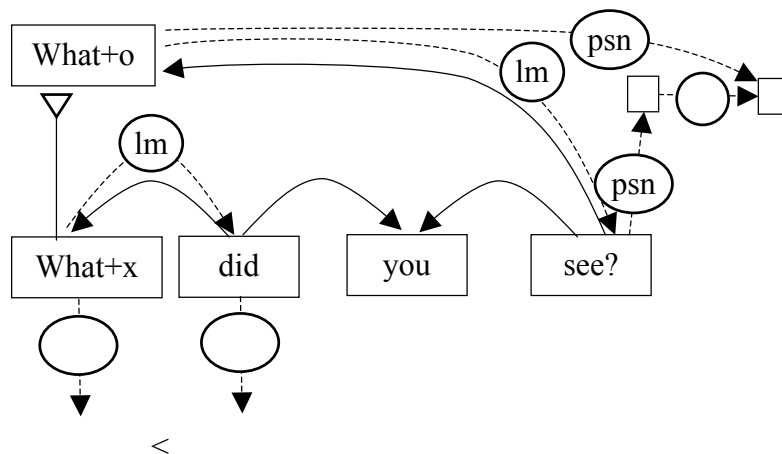


Figure 14: Extraction with landmarks, positions and overrides

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 we 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.

For example, non-configurational languages pose problems for standard versions of HPSG, as illustrated by this Wambaya sentence (Bender 2008):

(8) Ngaragana-nguja ngiy-a gujinganjanga-ni jiyawu
 having-grog.ACC 3sg.past mother give
 ngabulu.
 milk.ACC

‘(His) mother gave (him) milk with grog in it.’

The simplified literal gloss shows that both ‘having-grog’ and ‘milk’ are marked as accusative, which is enough to allow one to modify the other 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. Such freedom of order is easily accommodated if landmarks are independent of dependencies: the auxiliary verb is the root of the clause (as in English), and also the landmark for every word that depends on it, directly or 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 15, with dotted arrows again showing landmark relations.

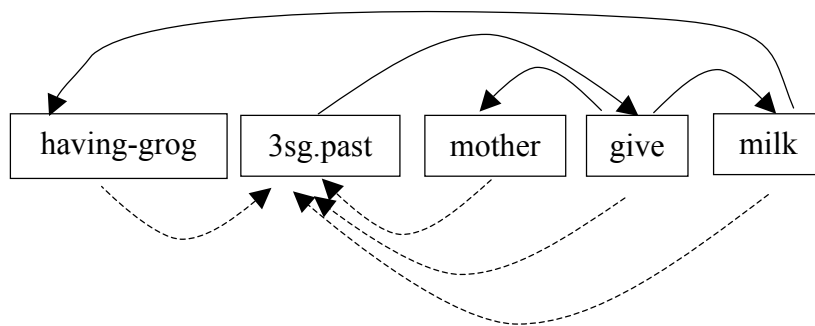


Figure 15: A non-configurational structure

The same formal apparatus allows an analysis of pied-piping constructions which avoids the **semantically-problematic feature percolation of HPSG analyses** (Hudson 2017).

12. Mutual dependency

*Why is it
semantically
problematic?
HPSG-interested readers
may want to know.*

And finally we must return briefly to the question that exercised the minds of the Arabic grammarians in the Abbasid Caliphate: is mutual dependency possible? For plain-vanilla PS and DS the answer has to be no for the simple reason that the notation uses the vertical dimension to represent dominance, and geometry doesn't allow two nodes each to be higher than the other. But for both HPSG and WG the answer is yes, and can be illustrated with the very simple example *What happened?* In HPSG *what* dominates *happened*, but its trace is dominated by *happened* (Müller 2018, 300). In WG, *happened* is the complement of *what*, but *what* is the subject of *happened*, so each depends on the other.

Combining this fact with the earlier discussion of word order, the mutual dependency has consequences for extraction: extraction possibilities differ between main and subordinate clauses.

(9) Then what happened?

(10) *I wonder then what happened.

The WG explanation (Hudson 2003) is that the main clause allows *happened* to be the landmark of both *then* and *what*, but the subordinate clause forces *what* to be the landmark of *happened*, which means that *then* would be separated from *happened* by a word with a higher landmark, giving the illegal tangling circled in Figure 16. I don't know whether HPSG allows a similar explanation for this rather curious restriction.

allow

There are two daughters one

Containing what

and the other a SVP consisting of V and a trace.

So what does not dominate happened.

I am not sure whether subjects are extracted at all in subj-wh sentences.

You or we or somebody should find out before publication.

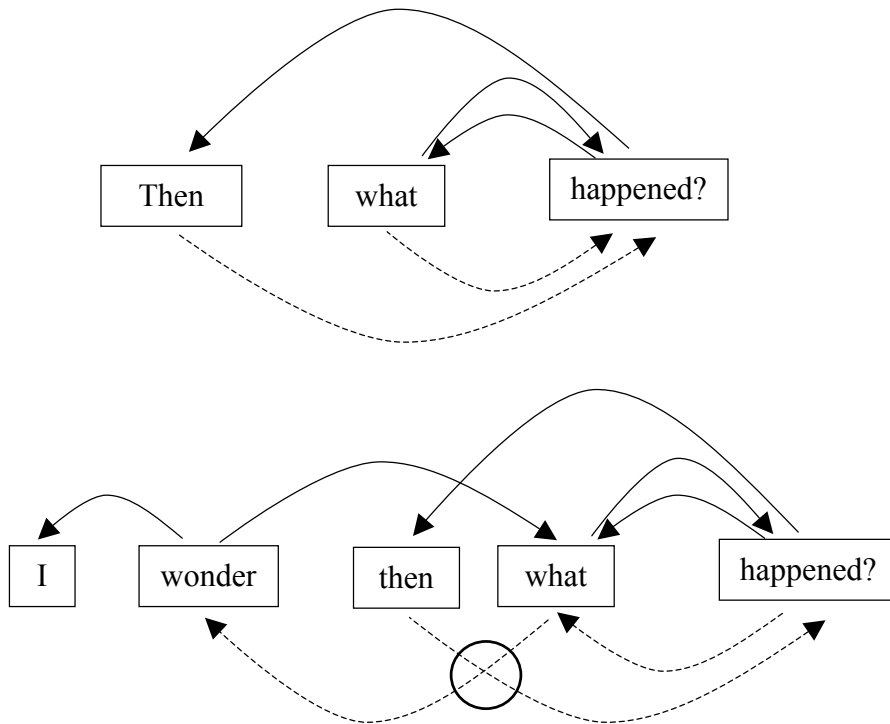


Figure 16: A problem caused by mutual dependency

13. HPSG without PS?

What is the problem?
Could you mention it in the caption?

This chapter raises a fundamental question for HPSG: does it really need PS? Introductory textbooks present PS as an obvious and established approach to syntax, but the early sections of this chapter 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. The challenge for HPSG, then, is to explain why PS is a better basis than DS. This may be possible but the debate has hardly started. The only attempt I am aware of is the excellent and compelling discussion of Welsh mutation (Tallerman 2009) which shows that phrase boundaries are needed; but a small extension of DS could add phrase boundaries without embracing the whole of PS, so this remains an unexplored research area for the HPSG community.

But suppose the debate favoured DS. Would that be the end of HPSG? Far from it. The analysis of words could remain as it is, and the only change needed would be the replacement of phrasal boxes by a single list of words. Here again is the example with which we started:

(11) Smart students enjoy syntax.

We have seen that the result of combining a word with a dependent is a distinct sub-token of the same word, so the basic word *students* is distinct from *students+a*, a new token whose dependent is the adjunct *smart* and which means 'smart students'. Similarly, combining *enjoy* with its object and subject produces two sub-tokens, *enjoy+o* and *enjoy+s*, meaning respectively 'enjoy syntax' and 'smart students enjoy syntax'. Each of these collections of tokens and sub-tokens comprises an unordered set such as $\{students, students+a\}$, but the basic tokens are (of course) an ordered set (distinguished as such by round brackets). The sentence would therefore have the following basic structure:

(12) (*smart*, $\{students, students+a\}$, $\{enjoy, enjoy+o, enjoy+s\}$, *syntax*)

Each word in the analysis stands for a whole box of attributes which include isa links between tokens and sub-tokens, and also syntactic dependency links to other words in the set. The internal structure of the boxes would look very much like

maybe
more
detail here

standard HPSG AVMs, so the only change would be in the relations among the boxes.

References

Barrett, Solomon. 1845. *The Principles of Grammar: Being a Compendious Treatise on the Languages, English, Latin, Greek, German, Spanish, and French. Founded on the Immutable Principle of the Relation Which One Word Sustains to Another.* (Revised Edition 1857). Cambridge: Metcalf.

Becker, Karl Ferdinand. 1831. *Schulgrammatik Der Deutschen Sprache.* Prague.

Bender, Emily. 2008. "Radical Non-Configurationality without Shuffle Operators: An Analysis of Wambaya." In *Proceedings of the 15th International Conference on Head-Driven Phrase Structure Grammar*, edited by Stefan Müller, 6–24. CSL! Publications.

Billroth, Johann. 1832. *Lateinische Syntax Für Die Obern Klassen Gelehrter Schulen.* Leipzig: Weidmann.

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

Blevins, James, and Ivan Sag. 2013. "Phrase Structure Grammar." In *The Cambridge Handbook of Generative Syntax*, edited by Marcel den Dikken, 202–25.

Bresnan, Joan. 2001. *Lexical-Functional Syntax.* Oxford: Blackwell.

Buszkowski, W. 1989. "Principles of Categorical Grammar in the Light of Current Formalisms." In *The Vienna Circle and the Lvov-Warsaw School.* Nijhoff International Philosophy Series 38. Dordrecht: Springer.

Chen, Danqi, and Christopher Manning. 2014. "A Fast and Accurate Dependency Parser Using Neural Networks." In *Proceedings of EMNLP (Empirical Methods on Natural Language Processing) 2014*, edited by Alessandro Moschitti, Bo Pang, and Walter Daelemans.

Chomsky, Noam. 1957. *Syntactic Structures.* The Hague: Mouton.

———. 1970. "Remarks on Nominalizations." In *Readings in Transformational Grammar*, edited by R Jacobs and Peter Rosenbaum,

184–221. Waltham, MA: Ginn.

Dahl, Östen. 1980. "Some Arguments for Higher Nodes in Syntax: A Reply to Hudson's 'Constituency and Dependency'." *Linguistics* 18: 485–88.

Dik, Simon. 1989. *The Theory of Functional Grammar: The Structure of the Clause*. Dordrecht: Foris.

Eppler, Eva. 2004. "The Syntax of German-English Code-Switching." UCL.

Frege, Karl Alexander. 1850. *Hilfsbuch Zum Unterrichte in Der Deutschen Sprache Für Kinder von 9 Bis 13 Jahren* (Second Edition). Wismar.

Gisborne, Nikolas. 1996. "English Perception Verbs." UCL, London.

Halliday, Michael. 1961. "Categories of the Theory of Grammar." *Word* 17: 241–92.

———. 1967. "Notes on Transitivity and Theme in English. Part 2." *Journal of Linguistics* 3: 199–244.

———. 1985. *An Introduction to Functional Grammar*. London: Arnold.

Hays, David. 1964. "Dependency Theory: A Formalism and Some Observations." *Language* 40: 511–25.

Hockett, Charles. 1958. *A Course in Modern Linguistics*. New York: Macmillan.

Hodges, Wilfrid. 2005. "A Context Principle." In *Intensionality*, edited by Reinhard Kahle, 42–59. Wellesley Mass.: Association for Symbolic Logic and A. K. Peters.

Hudson, Richard. 1971. *English Complex Sentences. An Introduction to Systemic Grammar*. Amsterdam: North Holland.

———. 1976. *Arguments for a Non-Transformational Grammar*. Chicago: Chicago University Press.

———. 1984. *Word Grammar*. Oxford: Blackwell.

———. 1990. *English Word Grammar*. Oxford: Blackwell.

———. 2003. "Trouble on the Left Periphery." *Lingua* 113: 607–42.



small
letter

? Dissertation?

———. 2007. *Language Networks: The New Word Grammar*. Oxford: Oxford University Press.

———. 2010. *An Introduction to Word Grammar*. Cambridge: Cambridge University Press.

———. 2017. “**Pied Piping in Cognition**.” *Journal of Linguistics* Online, 1–54. <https://doi.org/10.1017/S0022226717000056>.

Imrényi, András. 2013. “Constituency or Dependency? Notes on Sámuel Brassai’s Syntactic Model of Hungarian.” In *VLxx. Papers Presented to László Varga on His 70th Birthday*, by Péter Szigetvári, 167–82. Budapest: Tinta.

Jespersen, Otto. 1937. *Analytic Syntax*. Chicago: University of Chicago Press.

Kahane, Sylvain. forthcoming. “How Dependency Syntax Appeared in the French Encyclopedia: From Buffier (1709) to Beauzée (1765).” In *The Origins of Dependency Syntax. A Historical Survey from Panini to Tesnière*, edited by András Imrényi and Nicolas Mazziotta. Amsterdam: John Benjamins.

Keenan, Edward. 1976. “Subject and Topic.” In *Towards a Universal Definition of “Subject”*, edited by Charles Li, 303–33. New York: Academic Press.

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, and Joakim Nivre. 2009. “Dependency Parsing.” *Synthesis Lectures on Human Language Technologies* 2: 1–127.

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

Luhtala, A. 1994. “Grammar, Early Mediaeval.” In *Encyclopedia of Language and Linguistics*, edited by Ronald Asher, 1460–68. Oxford: Pergamon.

Marneffe, Marie-Catherine de, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher Manning. 2014. “Universal Stanford Dependencies: A Cross-Linguistic

- Typology.” In Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)., 4585–92.
- Matthews, Peter. 1993. *Grammatical Theory in the United States from Bloomfield to Chomsky*. Cambridge: Cambridge University Press.
- Mel’cuk, Igor. 1988. *Dependency Syntax : Theory and Practice*. Albany: State University Press of New York.
- Morell, John Daniel. 1855. *The Analysis of Sentences Explained and Systematised, after Beckers’ German Grammar*. Fifth Edition. London: Robert Theobald.
- Müller, Stefan. 2018. *Grammatical Theory: From Transformational Grammar to Constraint-Based Approaches*. Second Revised and Extended Edition. Textbooks in Language Sciences 1. Berlin: Language Science Press.
- Nesfield, J.C. 1898. *Outline of English Grammar*. In Five Parts. London: Macmillan.
- Nida, Eugene. 1943. “A Synopsis of English Syntax.” UCLA.
- Nivre, Joakim. 2006. “Constraints on Non-Projective Dependency Parsing.” In *Eleventh Conference of the European Chapter of the Association for Computational Linguistics*.
- Owens, Jonathan. 1988. *The Foundations of Grammar: An Introduction to Mediaeval Arabic Grammatical Theory*. Amsterdam: Benjamins.
- Percival, Keith. 1976. “On the Historical Source of Immediate Constituent Analysis.” In *Notes from the Linguistic Underground*., edited by James McCawley, 229–42. London: Academic Press.
- Perlmutter, David, and Paul Postal. 1983. “Some Proposed Laws of Basic Clause Structure.” In *Studies in Relational Grammar 1*, 81–128. Chicago: 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, and Ivan Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: Chicago University Press.
- Postal, Paul. 1964. *Constituent Structure. A Study of Contemporary*

Models of Syntactic Description. Supplement to IJAL 30.1.

Reed, Alonzo, and 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. (1890 Edition). New York: Clark and Maynard.

Robins, Robert. 1967. A Short History of Linguistics. London: Longman.

Robinson, Jane. 1970. "Dependency Structure and Transformational Rules." *Language* 46: 259–85.

Seguin, Jean-Pierre. 1993. L'invention de La Phrase Au XVIIIe Siècle : Contribution à l'histoire Du Sentiment Linguistique Français. Louvain and Paris: Peeters and Société pour l'information grammaticale.

Sériot, Patrick. 2004. "L'affaire Du Petit Drame : Filiation Franco-Russe Ou Communauté de Pensée? (Tesnière et Dmitrievskij)." *Slavica Occitania*, Toulouse, 93-118, .

Sgall, Petr, Eva Hajicová, and Jarmila Panevova. 1986. The Meaning of the Sentence in Its Semantic and Pragmatic Aspects. Pro Functional-Generative Description. Prague: Academia.

Siewierska, Anna. 1991. Functional Grammar. London: Routledge.

Tallerman, Maggie. 2009. "Phrase Structure vs. Dependency: The Analysis of Welsh Syntactic Soft Mutation." *Journal of Linguistics* 45: 167–201.

Tesnière, Lucien. 1959. *Éléments de Syntaxe Structurale*. Paris: Klincksieck.

———. 2015. *Elements of Structural Syntax*. Translated by Timothy Osborne and Sylvain Kahane. Benjamins.

Zwicky, Arnold. 1985. "The Case against Plain Vanilla Syntax." *Studies in the Linguistic Sciences* 15: 1–21.