



# Notational Variants and Cognition: The Case of Dependency Grammar

Ryan M. Nefdt<sup>1</sup> · Giosué Baggio<sup>2</sup>

Received: 29 September 2021 / Accepted: 23 December 2022  
© The Author(s) 2023

## Abstract

In recent years, dependency grammars have established themselves as valuable tools in theoretical and computational linguistics. To many linguists, dependency grammars and the more standard constituency-based formalisms are notational variants. We argue that, beyond considerations of formal equivalence, cognition may also serve as a background for a genuine comparison between these different views of syntax. In this paper, we review and evaluate some of the most common arguments and evidence employed to advocate for the cognitive or neural reality of dependency grammars in linguistics, psycholinguistics, or neurolinguistics. We then raise the possibility that the abilities to represent and track, alternatively or in parallel, constituency and dependency structures co-exist in human cognition and are constitutive of syntactic competence.

## 1 Introduction

Dependency grammar (DG) has become ubiquitous as a syntactic formalism in applied linguistic settings, including natural language processing (NLP), psycholinguistics, and neurolinguistics. Yet, rarely do practitioners in these fields consider whether DG may be more than a theoretical convenience and whether it is endowed with any cognitive or neural plausibility or reality. Furthermore, many scholars in formal and theoretical linguistics assume that dependency formalisms are ‘notational variants’ of the more standard phrase structure or constituency-based formalisms (Chomsky, 1972; 2000a; 2000b; Johnson, 2015). In this paper, we will reassess the main arguments why dependency grammars may be preferable to other formalisms, in particular to constituency-based grammars. We accept that, from a formal perspective, these

---

✉ Ryan M. Nefdt  
Ryan.Nefdt@uct.ac.za

<sup>1</sup> Department of Philosophy, University of the Cape Town, Cape Town, South Africa

<sup>2</sup> Department of Language and Literature, Norwegian University of Science and Technology, Trondheim, Norway

frameworks may indeed be hard to distinguish. However, we argue that cognition might prove to be a useful deciding factor in the appreciation of what unique contributions dependency analyses make to the understanding of human syntactic abilities.

In Sect. 2, we briefly introduce dependency grammar and we contrast some of its core features with those of phrase structure grammars. In Sect. 3, we describe the ‘received position’, the concept of ‘notational variance’, and some formal theory issues. In Sect. 4, we assess whether linguistic arguments fare better than formal language theory in deciding between dependency grammar and phrase structure. In Sect. 5, we consider further theoretical reasons for favoring dependency analyses, and we conclude that they are not any more decisive than the cross-linguistic data. In Sect. 6, we evaluate some of the recent literature in psycholinguistics and cognitive neuroscience and make a case for a parallel picture of dependency and constituency within human syntactic cognition. Lastly, in Sect. 6.4, we compare aspects of our view to the work done in Lexical Functional Grammar (LFG), before concluding.

## 2 What is Dependency Grammar?

Let us begin by defining dependencies and characterizing certain basic properties of dependency grammars. Essentially, dependencies are binary, asymmetric governance relations which hold between words in string sets. If word *A* dominates or *governs* word *B*, then word *B* *depends* on word *A*. Accordingly, word *A* is called a ‘head’ and word *B* a ‘dependent’. These relations are usually labelled in dependency trees, which are modelled using rooted directed acyclic graphs (see Chapter 14, Jurafsky & Martin 2021). For anything modelled on this sort of structure, one vertex acts as the root, from which directed edges or arcs are connected to other vertices. The edges are directed, in that they form ordered pairs of vertices; they are acyclic, in that there is no route back from one vertex to another: cycles or ‘loops’ are essentially banned. In dependency grammar, the vertices are words and the edges are the dependency arcs connecting them to one another and to the root in a sentence, usually the main verb. In brief, a “dependency tree representation of syntactic structure emphasizes the functional role of a word in a sentence” (de Marneffe & Nivre, 2019: 198).

Dependency trees look “flat” compared to the hierarchical syntactic structures of traditional phrase structure grammars, in which nested constituency plays a greater role. Compare Figs. 1 and 2 below:

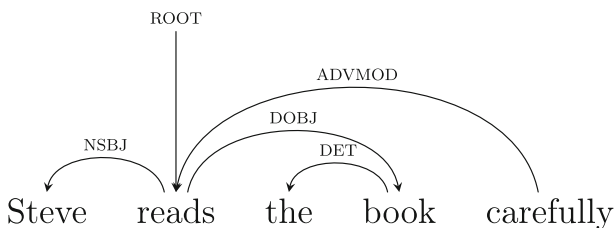
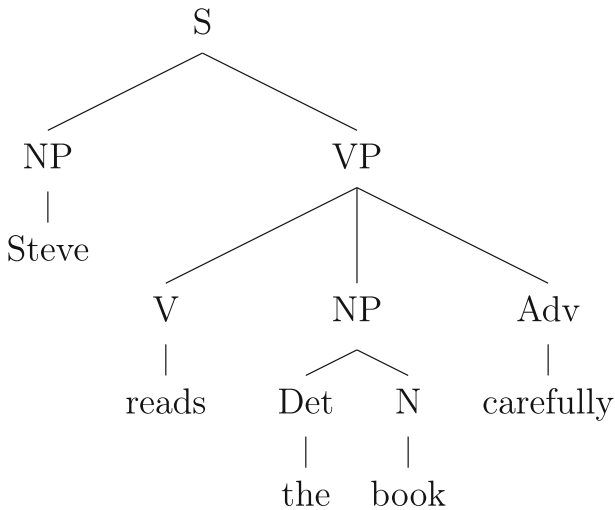


Fig. 1 Dependency graph for *Steve reads the book carefully*.



**Fig. 2** Phrase structure tree for *Steve reads the book carefully.*

The arrows or edges in Fig. 1 indicate the direction of the dependency (no hierarchy is intentionally represented). The verb ‘reads’ does not depend on anything else in the sentence, it is the *root*, while every other word does depend on something else, not necessarily adjacent to it.

Consider now the phrase structure tree of the same sentence in Fig. 2. Here, the structure comprises hierarchical constituency and phrases ordered by relations on trees, such as c-command (Reinhart, 1976; Chomsky, 1981).<sup>1</sup> A node X c-commands a node Y if the first branching node that dominates X also dominates Y, and if X does not dominate Y and Y does not dominate X. The phrase structure representation contains non-terminals, while dependency graphs do not. One element can govern multiple dependents directly in dependency grammar, which is not the case in the standard use of phrase structure, i.e., X-bar theory (unlike the tree in Fig. 2).

Despite the general shape of dependency analysis, dependency grammars are best seen as a class of formalisms sharing a family resemblance: no two frameworks may exactly correspond in their labelling practices, in their chosen dependency relations, and in decisions on non-obvious cases.<sup>2</sup> For instance, Word Grammar (WG) provides a ‘cognitive approach’ based on default inheritance hierarchies and on network links (Hudson, 1990; 2007), whereas Meaning-Text Theory (MTT) posits seven ‘strata’ of representation and ordering rules for discontinuities (Mel’cuk, 1988, 2011).

There are, however, tenets from which all dependency analyses generally proceed. These are relevant to both DG’s alleged methodological advantages (Sects. 4 and 5) and to its psychological and neural plausibility or reality (Sect. 6). These are: (a) a key role of individual words, i.e., ‘lexicalization’; (b) a binary, unidirectional dependence relation between words; and (c) a single ‘layer’ of syntactic representation. In terms

<sup>1</sup> Usually in formalisms such as X-bar theory, only binary branching is allowed, see Jackendoff (1977).

<sup>2</sup> This situation in part has led to the pursuit of Universal Dependencies (UD), which target a unified annotation and analysis for cross-linguistic data. See [universaldependencies.org/introduction.html](http://universaldependencies.org/introduction.html)

of (a), the lexicalization of a grammar involves associating each elementary structure with a lexical item or terminal node. Concretely, Rambow and Joshi (1997: 172) state that a grammar is lexicalized when “every elementary structure is associated with exactly one lexical item, and if every lexical item of the language is associated with a finite set of elementary structures in the grammar”. Importantly, Debusmann and Kuhlmann (2000: 371) also note that, “[i]f the underlying grammar is lexicalized, then there is a one-to-one correspondence between the nodes in the derivation tree and the positions in the derived string: each occurrence of a production participating in the derivation contributes exactly one terminal symbol to this string”. Dependency grammars differ from constituency grammars also in that the former are lexicalized from the outset: the elementary structures are the lexical items, or more precisely, nodes labeled with terminal elements (i.e., anchors). Dependency relations can also be posited for multiple ‘layers’ of linguistic form: phonological, syntactic, and semantic. This approach is called ‘multi-stratal’ DG in the literature (Sgall et al., 1986). Here, we focus on ‘mono-stratal’ versions, where syntactic structure is primarily represented.

This way of doing syntax shares certain points with pre-theoretical insights into grammar and was originally conceived of by Tesnière (1959), who initially regarded it as a cognitive approach to grammar with the goal of representing how the “mind perceives” syntactic connections between words.<sup>3</sup> This theme is further developed by Hudson (1990; 2007), who argues for a similar psychological reality of dependency structures in his Word Grammar. Despite the alternative origins of this grammatical formalism, many linguists still consider dependency grammar a ‘notational variant’ of the phrase structure approach to syntax. In the next section, we examine this position and thereafter attempt to tease the formalisms apart *via* conventional methods before suggesting that cognition might be the only viable realm of contrast and comparison.

### 3 The Received Position

The relationship between dependency grammar and phrase structure grammars has been a long-standing problem in linguistics.<sup>4</sup> Two early formal results aimed to settle the matter: those of Gaifman (1961) and Hays (1961, 1964). To appreciate these results we need to briefly state two important criteria for comparing grammar formalisms in formal language theory, namely weak and strong generative capacity.

The standard definitions of these terms have been held relatively stable since their introduction in Chomsky (1963). Therein, Chomsky states that “a grammar *weakly*

<sup>3</sup> Tesnière (1959) presents the gist of early DG as follows: “The sentence is an organised whole; its constituent parts are the words. Every word that functions as part of a sentence is no longer isolated as in the dictionary: the mind perceives connections between the word and its neighbours; the totality of these connections forms the scaffolding of the sentence. The structural connections establish relations of dependency among the words. Each such connection in principle links a superior term and an inferior term. The superior term receives the name *governor* (*régissant*); the inferior term receives the name *dependent* (*subordonné*)”.

<sup>4</sup> For the interpretation of relevant formal results in this section, we will be largely assuming that we are in the realm of context-free grammars. There is much controversy as to whether phrase structure grammars fall within this class or even what the precise definition of the term is. Manaster-Ramer and Kac (1990) offer a systematic overview of the ‘many meanings’ of the term in linguistics.

generates a set of sentences” and “*strongly generates* a set of structural descriptions” (1963: 60). From this, weak generative capacity can be characterized in terms of the class of formal languages generated by a specific grammar in the Chomsky Hierarchy. For instance, context-free grammars (CFGs) generate the set of context-free languages (CFLs), regular grammars generate the set of regular languages, and so on. Peters and Ritchie (1973) showed that the transformational grammar of *Aspects* generates the set of recursively enumerable languages: it is a Type-0 or unrestricted grammar.<sup>5</sup>

Two grammars, then, are said to be *weakly equivalent* if they generate or license the same sets of strings. If for a grammar of the first type of formalism there exists a grammar of the second type that generates or licenses the same sets of strings, then the two grammars are said to be weakly equivalent. Gaifman’s (1961) result establishes this correspondence between dependency grammar and phrase structure grammar. Specifically, he proved that dependency grammars are a special case of context-free grammars and a proper subset of phrase structure grammars, that is, the ‘immediate constituency grammar’ of the time. As Chomsky (1965) noted, however, “discussion of weak generative capacity marks only a very early and primitive stage of the study of generative grammar. Questions of real linguistic interest arise only when strong generative capacity (descriptive adequacy) and, more important, explanatory adequacy become the focus of discussion” (61).

Strong generative capacity involves a particular kind of structural equivalence or ‘equivalence of structural descriptions’ relevant for deeper linguistic questions related to language acquisition. Two grammars are *strongly equivalent* if they assign the same structural descriptions to the same strings. Here, Gaifman only shows one direction of equivalence, from dependency structures to phrase structures, but *not vice versa*. As Hays (1964: 522) states, “the two theories are not strongly equipotent”, though he suggests that this should count in favor of dependency grammar, as the properties of phrase structure it fails to mimic are unlikely to have ‘linguistic applications’. One specific property in question is proven in Hays (1961). If the phrase structure grammar licenses an infinite set or allows recursive categories (a “[phrase structure] system of infinite degree”, in his terminology), then a simple correspondence with a dependency grammar is impossible. Yet, the jury is still out on a precise characterization of strong generative capacity and therefore of the equivalence it involves. For instance, Levelt (1974) and Kuroda (1976) provided notable criticisms of the notion assessing its triviality, and Miller (1999) developed a model-theoretic treatment of the concept of strong generative capacity and its equivalence.

Despite a lack of consensus on formal aspects of the correspondence between DG and phrase structure grammars, linguists often still assume that they are equivalent as notational variants. Chomsky defines notational variance in the following manner:

Given alternative formulations of a theory of grammar, one must first seek to determine how they differ in their empirical consequences, and then try to find ways to compare them in the area of difference. It is easy to be misled into

<sup>5</sup> This means that all formal languages recognizable by a Turing machine can be generated by such grammars. See Berwick (1984) for reflection on whether the same holds for the subsequent Government and Binding theory, which eschewed the strong deletion rules and many individual transformations of its predecessor and incorporated a more parsimonious trace theory.

assuming that differently formulated theories actually do differ in empirical consequences, when in fact they are intertranslatable—in a sense, mere notational variants. (1972: 69)

If two grammars are notational variants, they have (possibly among other properties in common) the same empirical consequences and are equally empirically adequate. Exactly what this means in terms of weak and strong generative capacity is unclear. For if empirical adequacy is defined in terms of descriptive adequacy (and *a fortiori* observational adequacy) à la Chomsky (1965), then dependency and phrase structure grammars must have the same empirical consequences, since they generate the same sets of sentences. Presumably, these sets would be those identified by native speaker intuitions or judgements, according to the definition of descriptive adequacy. On the other hand, if empirical adequacy also requires explanatory adequacy, as is suggested by Chomsky's quote above, that is, an account of language acquisition, then something stronger, such as structural equivalence or strong generative capacity, may be needed. Johnson (2015) applies a measure-theoretic account of notational variance using the Merge operation of Minimalism (Chomsky, 1995). He then argues that, to the extent that there is structural overlap between grammars, there lies the empirical fruitfulness of the theory. He makes use of an analogy from physics between symmetries which are invariant under certain transformations and formal grammars. Thus, he attempts to liberate the concept of equivalence from its usual negative connotations by accepting and exploiting the idea that grammar formalisms, such as dependency grammar and phrase structure grammar, are indeed equivalent. Johnson wants to see his proposal as addressing Quine's challenge concerning the 'psychological reality' of equivalent grammar formalisms: if X and Y are weakly equivalent, or 'behaviourally equivalent' in Quine's terms, which one is realised in the human mind and brain? The answer is that both are, according to Johnson. This is no doubt an intriguing proposal, but it is unclear what leverage it has for linguistic analysis and language acquisition. Relatedly, Stabler (2019) has argued that different mathematical foundations for syntax, present in phrase structure grammars, in dependency grammars, and in optimality theoretic formalisms, have common roots in formal tree languages. Stabler's formal arguments for the insufficiency of each individual framework, together with their joint relevance for modeling linguistic competence, dovetail with the conclusion we reach in Sect. 6 on more cognitive grounds.

The situation we are left with is one in which the received position largely has its roots in early research in formal language theory and in mathematical comparisons between the different grammars. In some cases, the alleged equivalence is based on weak generative capacity;<sup>6</sup> in other cases, the notion of strong generative capacity is either formally or informally treated. Arguably, mathematics is not going to settle the question of the equivalence of the two formalisms under discussion. As Rambow and Joshi (1997: 167) note of the specific comparison in the current work:

<sup>6</sup> There are proofs of the weak equivalence of various grammar formalisms, such as Dependency Grammar and versions of Phrase Structure Grammar (PSG), Tree-Substitution Grammar (TSG), and Categorical Grammar (CG), among others (Vijay-Shanker & Weir, 1994).

A mathematical comparison between the underlying formal systems will [...] not tell us much about the linguistics of the two theories. Instead, we must ask how the formalisms affect the linguistic theories which are expressed in them.

It is on to this task that we now move. In the following section, we will try to show that linguistic evidence can provide insights into theory comparison, but it ultimately falls short of deciding on the equivalence of DG and constituency-based formalisms. This task is better suited for cognitive science, we will argue in Sect. 6.

## 4 Linguistic Arguments in Favour

### 4.1 Cross-Linguistic Coverage

From a linguistic perspective, empirical adequacy, understood as cross-linguistic data coverage, is of paramount importance for theories of grammar. Simplifying, if some formalism explains or models more of the data, then it is preferable *ceteris paribus* to a more limited competitor.<sup>7</sup> For example, tree-adjointing grammars (TAGs) encompass the results of tree-substitution grammars (TSGs), but have the added bonus of dealing more effectively with linguistic phenomena like adjectival modification (via adjunction) in languages like English. In this sense, TAGs seem preferable on grounds of empirical adequacy. Below, we evaluate some standard arguments for dependency grammars in terms of cross-linguistic coverage. We will show that, although the DG formalism is able to model free-word order languages and long distance dependencies (LDDs) fairly naturally, it still suffers from sufficiency issues, and where it is truly empirically advantageous, it suffers from complexity issues as well.

#### 4.1.1 Free-Word Order Languages

The majority of the world's extant languages exhibit either SOV or SVO patterns in declarative sentence structures (Greenberg, 1963; Givón, 1975; Bickerton, 1981). The evidence even suggests that, if there ever was an ancestral language, or Proto-human, from which all other languages were derived, then it had SOV order, which, through diffusion, resulted in SVO or some other word order in subsequent languages in the phylogenetic tree (Gell-Mann & Ruhlen, 2011). In reality, many languages incorporate a mixed word order with a favored option or a default structure, while others, such as German, exhibit a primary alternation between SOV and SVO, for example.<sup>8</sup> Free word order is a phenomenon attested when the balance between different word order patterns is relatively equal, or when the frequencies of two or more word orders is similar in the given language. Inflected languages, such as Latin, Greek, or Russian, naturally tend to allow for more flexibility on this measure than do *more* isolating languages, such as Mandarin Chinese, English, or Afrikaans. In several of these languages, word order would still serve pragmatic functions. However, some of the former languages

<sup>7</sup> This is a simplification because data coverage is not the only or main goal of theory. Computational complexity, tractability, learnability etc., also act as constraints on theory (van Rooij & Baggio, 2021).

<sup>8</sup> See Ferrer-I-Cancho (2016b) for a formal framework for word order variation and alternation.

show signs of phenomena such as ‘word order freezing’, where some configurations are interpretable exclusively in a specific sequence, despite case markings allowing for other possible readings (Jacobson, 1958; Zeevat, 2006).

Phrase structure grammars are notoriously sensitive to word order in a sentence. In contrast, a single dependency tree can represent two or more differently ordered sentences. As de Marneffe and Nivre (2019: 205) conclude, “dependency trees can thus capture generalizations better in languages with free or flexible word order than phrase structure trees”. This is also evidenced by the number of treebanks favoring dependencies in their annotations (Bouma et al., 2000; Kromann, 2003; Oflazer et al., 2003), and by typological work comparing structures in different languages (Nichols, 1986; Liu, 2008; Futrell et al., 2015b; Chen & Gerdes, 2017; Ferrer-i-Cancho et al., 2022).

The typological point is not only one of parsimony (see Sect. 5): it is not merely the case that DG enables more compact analyses of freer word order languages than phrase structure grammars, but rather that simple phrase structure representations can miss similarities between structures. As noted by de Marneffe and Nivre (2019: 206), languages might use different constructions for nonverbal predication, e.g., a copula strategy, as English does, versus a ‘zero strategy’, as in Russian:

- (1) a. The houses are new.  
b. Doma novye.

The equally acceptable ‘novye doma’, in relation to (1b), is represented by the same dependency tree, if necessary using indices marking surface word order. In addition, “differences in strategy (...) are clearly demonstrated by a dependency representation, but would be harder to see in a phrase structure representation” (de Marneffe & Nivre, 2019: 206). Dependency grammar easily represents functional categories which are multiply realisable in some languages, a property it shares with other formalisms, like Lexical Functional Grammar.<sup>9</sup> Osborne (2014: 616) describes the situation as follows:

The necessity to place [one word] in front of the other [e.g., in ‘talk trash’ vs ‘trash talk’] means that constituency cannot produce tree structures that abstract away from linear order. Dependency is hence more capable of focusing on the one ordering dimension in isolation (on the vertical dimension). The discontinuities associated with free word order become less problematic because fewer crossing lines appear in the trees.

On the other hand, if word order freezing is a (strictly) syntactic phenomenon, then one might worry that some information might be lost by identical dependency trees for nonlicensed and licensed constructions alike, abstracting away from linear order. Still, the point remains that DG may be preferable from a cross-linguistic perspective, if coverage of world languages is the goal, which it is here. Formalisms with stricter word order constraints would be less desirable alternatives in such cases.<sup>10</sup>

<sup>9</sup> From a theoretical perspective, dependency grammar does not assume a universal syntactic level of representation that explains syntactic structure in all languages (e.g., UG in generative syntax), but does produce a uniform method of analysis applicable to many languages.

<sup>10</sup> It should be noted that unification-based grammar formalisms that incorporate feature structures and inheritance hierarchies, such as Head-driven phrase Structure Grammar (HPSG) and Sign-based Con-



### 4.1.2 Long-Distance Dependencies (Without Movement)

Another area in which dependency grammars may seem more empirically adequate than constituent-based grammars is in relation to long-distance dependencies (LDDs) in various languages. In many standard cases, relations between words hold in virtue of their proximity to one another. That is, properties of items in a sentence, such as words and morphemes, are determined by the properties of linearly adjacent items. For example, in the English noun phrase ‘an orange’, the shape of the indefinite article ‘an’ depends on the word immediately following it, here one beginning with a vowel. If a different word intervenes, then the form of the determiner would change based on that word, as in ‘a rotten orange’. By contrast, LDDs involve a kind of ‘action at a distance’, where dependencies hold between words not immediately adjacent to one another. *Wh*-sentences are canonical examples of such dependencies.

(2) What did Mary say about Tom?

In (2), ‘what’ depends syntactically on ‘say’, and not on the auxiliary ‘did’. In DG, this is the case for Universal Dependencies (UD) annotations. In some constituent-based theories, *movement* captures such ‘action at a distance’: here, ‘what’ moved from a lower position in the tree adjacent to the verb, leaving behind a trace or copy. This solution is part of a more general filler-gap strategy for dealing with constructions in transformational analyses where a sentence-initial filler is extracted and leaves a gap or *wh*-trace empty category. Dependency grammar has no need for movement operations or for filler-gap relations in the syntax. By the very design of dependency trees, there is, in a sense, nothing to explain in such cases. ‘Say’ is the DG graph root and the head upon which ‘what’ depends, as does the subject ‘Mary’. Nothing in the formalism demands adjacency for dependence: so, long-distance dependency comes for free. The flip-side of this is that dependency grammars have to find alternative ways of capturing important aspects of linearization, like fronting (see Müller, 2018).<sup>11</sup> According to certain axiomatizations of dependency grammar (Robinson, 1970) the following property is a condition on saturation of such structures:<sup>12</sup>

**Projectivity:** If A depends directly on B, and some element C intervenes between them (in the linear order of the string), then C depends directly on either A or B or some other intervening element (Debusmann, 2000: 4).

Projectivity is a way of characterizing syntactic dependencies between non-adjacent elements in a sentence. Not only does it deal effectively with *wh*-constructions, but also with anaphora, which may involve long-distance dependencies and agreement, such as of number and gender. This is even clearer in inflected languages with freer

---

struction Grammar (SBCG), may capture some of the strategies discussed here without sacrificing either hierarchy or constituency. Kahane (1996) provides an interpretation of HPSG as a dependency grammar, while Hudson (1990) deals with long-distance dependencies using the strategy of the older generalized phrase structure grammar (GPSG).

<sup>11</sup> We thank an anonymous referee for pointing this tension out.

<sup>12</sup> Along with (1) no word should depend on itself (not even transitively), (2) each word should have (at most) one governor, and (3) the dependency analysis should cover all words in a sentence. However, these conditions are not uniformly accepted across frameworks. Some analyses, like Hudson’s Word Grammar, allow for multiple governors/heads in cases of raising, for instance.

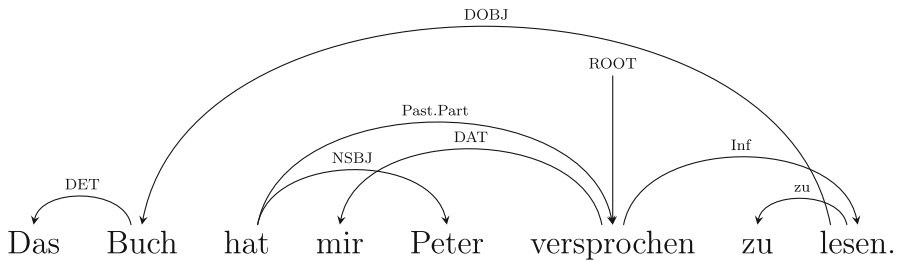


Fig. 3 Non-projective dependencies.

word order. Dependency grammar captures a wide range of syntactic constructions without the need for additional formal mechanisms such as movement, thus extending its cross-linguistic coverage to languages in which such constructions are ubiquitous. Yet, projectivity may come at a coverage cost. The condition itself formally amounts to a ban on crossing edges. The problem is that certain languages appear to require crossing edges for their syntactic analysis, in particular languages with relatively free word order. Consider the German sentence below and in Fig. 3 (Duchier, 2000):

- (3) Das Buch hat mir Peter versprochen zu lesen.  
 The book has me Peter promised to read.  
 The book Peter has promised me to read.

The edge from the past participle ‘hat’ to the subject ‘Peter’ crosses with the edge from the verb ‘versprochen’ to its dative object ‘mir’. Crossing edges are one way of dealing with such constructions. Muller (2018) identifies a number of other strategies employed by dependency grammarians. For instance, “one could assume additional mechanisms that promote the dependency of an embedded head to a higher head in the structure” (372). Furthermore, from the work of Bresnan et al. (1982), Shieber (1985), and Culy (1985) follows that context-free grammars (CFGs) or even lexicalized versions of CFGs are not quite adequate for capturing cross-serial dependencies in Dutch and aspects of Swiss German syntax (and Bambara), respectively. Because projective dependency grammars are weakly equivalent to CFGs (and can be induced by lexicalised CFGs), a result proven by Hays (1964), they too are inadequate for describing certain linguistic phenomena.

However, non-projective dependency structures also lead to increased complexity and resultant parsing costs. In the parallel literature on discontinuous constituents in phrase structure grammars, theorists have proposed devices, such as scrambling for dealing with these kinds of phenomena. It has also been noted that, “[i]nsofar as computational linguistics is concerned, these rules [e.g., non-projective rules] are like any other rules involving transformations—they become unfeasible if one proceeds to implement efficient parsing algorithms” (Debusmann, 2000: 9). This statement is somewhat misleading. Parsing in unconstrained non-projective dependency grammar is an NP-hard problem, while parsing of certain *mildly* non-projective dependency grammars can be made in polynomial time as for certain context-sensitive grammars. But there has been work conducted on ‘mildly non-projective’ dependency pars-

ing, which hopes to retain the well-behaved nature of projective dependency while capturing certain aspects of free word order languages.<sup>13</sup> By giving up projectivity, dependency grammars would face similar issues to those faced by phrase structure grammars with relation to discontinuity of constituents in immediate constituent theory. Despite this worry, cross-linguistic data seem to support theories which minimize the distance between heads and their dependents. Current research on ‘dependency length minimization’ (DLM) indicates that language users tend to opt for structures which avoid LDDs in production and comprehension. Formal results on complexity dovetail with empirical data, indicating that speakers across languages prefer shorter dependency lengths between words. In a cross-linguistic corpus study by Futrell et al. (2015: 10336), two predictions were confirmed:

First, when a grammar of a language provides multiple ways to express an idea, language users will prefer the expression with the shortest dependency length. Second, grammars should facilitate the production of short dependencies by not enforcing word orders with long dependencies.

This is broadly compatible with the property of projectivity, which has been argued to stem from DLM (Ferrer-i-Cancho 2006; Gomez-Rodriguez et al., 2017). Futrell et al.’s analysis of DLM in 37 languages is modelled on dependency grammar: dependency length is defined in terms of DG. Their findings and those of others, for example of Gildea and Temperley (2010), who used English and German data, show that dependency lengths are generally shorter than random baselines in English and than optimal linearizations in German based on extraction and reordering of text (see also Ferrer-i-Cancho, 2004; Ferrer-i-Cancho et al., 2022). But cross-linguistic evidence does not vindicate DG. Also, DLM motivates the evolution of constituent structure too, which is based on the absolute minimum length between words, ‘length 1’ (Ferrer-i-Cancho, 2004; Futrell et al., 2015), or the property of linear adjacency. We return to these issues in Sect. 6.

## 5 Theoretical Arguments in Favour

We have seen how the nature of syntactic representations in dependency formalisms has often been touted as an advantage in cross-linguistic and typological studies. In this section, we delve into the claim that dependency grammar is a more convenient or useful tool for the study of language, both from a theoretical standpoint as well as from an applications perspective. This is another key task in the quest to address the received position about notational variance. Chomsky (2000b) makes a similar move when he suggests that the equivalence between the derivational and representational approaches to grammar is akin to the difference between  $25 = 5^2$  and  $5 = \sqrt{25}$ , while he nevertheless maintains that the derivational approach affords unique theoretical insights. Here, we focus on three main theses: (1) dependency grammars are simpler from a structural perspective than most other alternatives; (2) they involve more trans-

<sup>13</sup> See Pitler et al. (2013), Kuhlmann (2013), and Gómez-Rodríguez (2016). For some background on the complexity of incremental dependency parsing algorithms, see Nivre (2008).

parent semantics or argument structure; and (3) they serve applications in NLP and psycholinguistics well. We will show that, once again, there is a kernel of truth to all the above statements, but not without significant qualification.

## 5.1 Representational and Conceptual Simplicity

One alleged property of dependency grammar, often peddled by its practitioners, is its representational simplicity. Osborne (2014: 624) is explicit in this regard:

The number of nodes in dependency-based structures tends to be approximately half that of constituency-based structures. The greatly reduced number of nodes means a greatly reduced amount of syntactic structure in general. The minimalism of these reduced structures then permeates the entire theoretical apparatus. Most DGs draw attention to this fact, and, in so doing, they are appealing to Occam's Razor: the simpler theory is the better theory, other things being equal.

Not only are fewer nodes, no movement, and 'flatter' syntactic structures believed to make dependency grammars more parsimonious than phrase structure grammars, but the 'transparency' of the dependency formalism is also viewed as an advantage. What is meant by 'transparency' is something analogous to conceptual simplicity. As de Marneffe and Nivre (2019: 208) state, "the core structure of dependency trees, that is, binary relations between lexical elements forming a tree, is a conceptually simple representation". The idea that binary relations in the grammar hold between lexical items (words) is, moreover, accessible to non-linguists. Therefore, DG seems to have intuitive purchase and wider appeal, as well as simpler structure.

From the above, the criterion of simpler structure as such could be understood in terms of three related properties: (1) dependency grammars involve fewer structural points or nodes, (2) dependencies map one-to-one onto lexical items, i.e., words, and (3) grammatical relations are encoded intuitively. Let us unpack these claims.

The idea of fewer nodes (1) is relevant from a complexity perspective. One node per word allows for the implementation of more efficient parsers: the parser only has "to link existing nodes together and not to postulate new ones" (de Marneffe & Nivre, 2019: 208). Instead, non-terminal phrase-level nodes have to be inferred for parsing of phrase structures. The lack of non-terminal nodes in DG representations involves less covert structure in the syntax, and covert structure requires motivation. In essence, this feature dovetails with Chomsky's (1995) proposals regarding 'economy of derivation' and 'economy of representation'. DG incorporates both aspects while the Minimalist program itself arguably has favored the former over the latter.

In terms of (2), a mapping from nodes to words means that the grammar does not employ mechanisms like movement, or nodes which terminate in empty strings or empty categories. The explanatory burden of describing how surface forms may be derived from deeper structures as these has beset the generative program, running from the Standard Theory (1957), through the Extended Standard Theory (1970), to Government and Binding (1981) and Minimalism (1995).<sup>14</sup> This is not an issue for

<sup>14</sup> Even though Minimalism finally shed the 'deep structure' postulate of earlier versions.

dependency grammar, which takes a more direct route to grammatical forms from surface structures, via linking of words in terms of standard grammatical relations.

Lastly, (3) pertains to the kind of relations holding between function words and content words. Across DG formalisms, and across labelling conventions, grammatical functions are encoded as dependency arcs. Traditional roles, such as subject, object, and conjunct, are commonly accepted even in frameworks that make non-standard theoretical choices in terms of additional semantic layers and valencies, for example Meaning Text Theory. This is one of the reasons why Universal Dependencies (UD), the typological initiative to create consistently annotated dependency treebanks, is a plausible project. UD assumes there is a shared set of dependency relations that can be uniformly and abstractly represented.<sup>15</sup> Without overt structure and grammatical relations across frameworks, a project of that kind would be doomed from the outset.

Although the above arguments make a reasonably strong case for applications of dependency analyses against phrase structure, there is a puzzling preoccupation with mirroring constituency across dependency frameworks. In Sect. 1, we represented the head-dependent relationship simply in terms of ordered pairs of words. In reality, however, finer-grained groupings have been proposed. In Word Grammar, variously coordinated elements, like conjuncts, are represented by means of square brackets, demarcating word strings. Kahane's (1997) bubble tree formalism allows dependency relations to hold between sets of nodes. Other approaches, including Mel'cuk (1988) and Garde (1977), introduced 'fragments' and 'significant elements' respectively to capture constituent-like units. Some dependency theorists (O'Grady, 1998; Osborne, 2005; Osborne et al., 2012) have even put forward a new notion, that of 'catenae'. A catena is defined as a word or a combination of words continuous with respect to dominance. In terms of graph-theory, a catena is any tree or subtree of a tree, but the notion of subtree is not equally applicable to constituency and dependency structures: the word clusters that constitute subtrees of dependency trees are not subtrees in the corresponding constituency structures. Importantly, the words in a catena do not need to be contiguous in the linear ordering: catenae are not strings or substrings. Constituents are special cases of catenae, where subtrees are complete.

The catena unit is much more flexible than the constituent, and the argument has therefore been put forward that the catena is better suited to serve as the basic unit of syntactic (and morphosyntactic) analysis than the constituent (Osborne, 2014: 620).

Despite the advantages of dependency structures, suitable notions of constituency might still need to play a role for a syntactic theory to achieve empirical adequacy and completeness. We return to this point more extensively below. Suffice it to say here that similar arguments have been used to motivate frameworks that combine aspects of constituency and dependency (or equivalent) structures, such as LFG. Predictably, however, there are trade-offs and limitations as to what can be achieved via adjustments to dependency representations and associated formal devices, such that,

<sup>15</sup> Of course, there are still thorny issues to deal with, such as the analysis of function words, which in UD get subordinated to content words, unlike in most other dependency grammar formalisms. See Osborne and Gerdes (2019) for more on this particular discrepancy.

for example, some of the aforementioned advantages seem to fall away. As a result, in computational applications in NLP, simple dependencies are incorporated, rather than the more constituent-based modifications (Kübler et al., 2009).<sup>16</sup> Nevertheless, several of the virtues discussed above directly relate to why dependency structures are generally utilized in NLP applications and, increasingly, in psycholinguistic work that embraces computational linguistics methods and results. We will get back to this in Sect. 5.3. But first we consider one additional aspect of theoretical convenience which some believe could favor DG over constituency-based grammars, namely the transparency of the interface with semantic structure.

## 5.2 Argument Structure and Semantics

The claim that dependency relations can transparently illuminate or convey semantic structures is another complex issue. On the one hand, dependency structures reflect argument structure in an overt manner. On the other hand, the lack of constituency precisely presents unique problems for the development of compositional semantics, as it is usually understood (Heim & Kratzer, 1998).

In phrase structure grammars, syntactic categories are determined exclusively by distributional properties and grammaticality. This is partly due to the ‘autonomy of syntax’ thesis—one of the core tenets of generative linguistics. The general manner in which argument structure is represented in those frameworks is through theta roles, which specify selectional restrictions on verbs, such as agent, patient, theme etc. These roles are invisible to the syntactic representations and require additional tools, such as the Theta Criterion, applied to X-bar theory in the Extended Standard Theory (Jackendoff, 1977; Chomsky, 1981). Because the X-bar formalism overgenerates structured representations, the Theta Criterion reins in aberrant forms and interfaces with systems of interpretation (logical form, LF).

It is generally believed that “[d]ependency structures are able to capture argument structure, the part of a sentence’s structure that specifies ‘who did what to whom,’ succinctly and accurately” (Penn, 2012: 169). Argument structure is centered around predicates, which are mostly verbs (but may also be other parts of speech), and their arguments, so one can technically read that information off any standard dependency tree. For example, the arguments that a verb takes are found following the arrows stemming from the root. Consider the following example:

- (4) Mary gave the book to Steven.

From the dependency graph in Fig. 4 one can see that the verb or predicate ‘gave’ takes three arguments: the subject (NSBJ) ‘Mary’, the modifier (NMOD) ‘to Steven’, and the direct object (DOBJ) ‘the book’. These are the agent, the recipient, and the theme, respectively. These thematic labels could just be appended to the dependency structure, without altering the nodes or the dependency arcs. Nor is a separate theta

<sup>16</sup> Interestingly, the realization of the necessity of alternative frameworks, or a dual approach, goes both ways. Osborne et al. (2011) discuss a trend in Minimalism toward dependency analyses and even construct a version of Merge in terms of catenae.

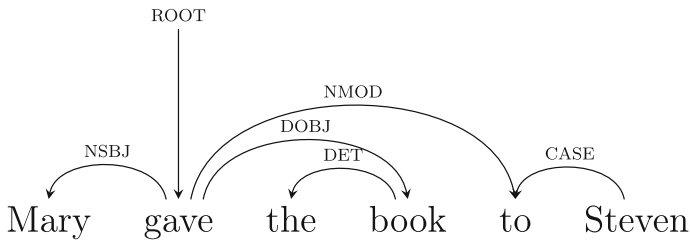


Fig. 4 Dependency graph for *Mary gave the book to Steven*.

grid necessary here. The argument structure is given in the dependency graph. This is highly compact and transparently captures semantic relations.

Despite the transparency of argument structure in dependency grammar, which is a distinctive advantage, dependencies have a theoretical drawback when it comes to compositional semantics. One standard approach in formal semantics constructs phrase and sentence meanings from constituent meanings and constituent structure by means of functional application. Heim and Kratzer (1998) call “Frege’s conjecture” the view that all meaning composition is the “saturation of an unsaturated meaning component”. Functional application in the lambda calculus is used to model formally this “saturation” process (Martin & Baggio, 2019). Each expression that participates in the composition process is assigned to an appropriate semantic type, these types can then be composed into more complex types depending on the syntactic construction underlying the expression. Functional application tracks one by one the operations required to build a phrase structure tree. This fact establishes a homomorphism between the syntactic algebra, based on phrase structure, and the semantic algebra, effectively creating a ‘rule-to-rule’ mapping in the traditional Montagovian picture. This one-to-one correspondence is the central methodological principle of formal semantics—the *principle of compositionality*, i.e., the idea that “the meaning of a complex expression is determined by the meanings of its constituents and by its structure” (Szabó, 2000: 1).<sup>17</sup> Functional application and the principle of compositionality presuppose viable notions of syntactic hierarchy and constituency, and it is just not clear how dependencies could figure into standard formal semantic accounts, following the methodology laid out by Heim and Kratzer (1998). Other formal issues arise in connection with compositionality, including the apparent clash between the asymmetric view of relations in DG (all syntactic dependency arcs are directed) versus the symmetric nature of some logico-syntactic relations. For example, coordinated structures like ‘John and Mary’ are symmetrical. However, dependency systems have to choose which word is the head: the UD framework chooses the left-most word as the head, here ‘John’, which is then linked to ‘Mary’ via a CONJ arc, and ‘Mary’ to ‘and’ via the CC (coordinating conjunction) arc. The clash as such is not problematic, but its immediate effects are: nothing in the meaning of conjunction corresponds to these asymmetric arcs, and that is a violation of ‘rule-to-rule’ compositionality.

<sup>17</sup> Some semantic theories, such as discourse representation theory (DRT) (Kamp, 1981), were initially deemed incompatible with compositionality. See Zeevat (1989) for a compositional treatment of DRT.

One alternative to standard interpretive semantics is Glue semantics, originally developed for Lexical functional grammar (Darymple et al. 2019). Glue semantics involves a linear logic for the composition of meaning. It is generally applicable since “[t]he only aspect of syntax that Glue presupposes is a notion of headedness, which is universal across formal syntactic theories” (Asudeh, 2022: 322). In fact, Glue has been proposed for dependency grammar in Bröker (2003) and Universal Dependencies in Gotham and Haug (2018). There are a number of components to Glue semantics. For instance, Glue is a resource logic which requires that all premises have to be used at most once in the semantic derivation. This is a departure from the classical logic assumed in standard truth-conditional semantics with types. Another aspect of linear logic is that it is commutative, which means that the semantics is blind to word order constraints. Asudeh (2022) attributes this to the Glue property of ‘autonomy of syntax’. The semantics is also somewhat independent and uses underspecification-like techniques, in which not all semantic ambiguities are derived from the syntax, as is the case in Montague Grammar. The semantics itself pairs expressions in a typed lambda calculus with those of a linear logic based on the syntax. This is what is called a ‘meaning constructor’.

Unfortunately, we do not have the space to enter into too many details here.<sup>18</sup>

### 5.3 Applications to NLP and Psycholinguistics

Dependency grammars are widely used in NLP applications, including in web search, where dependency arcs may be preferred to regular terms, in question answering, in sentiment analysis etc. (Eisenstein, 2019), but the motivations for the prevalence of dependency formalisms, relative to constituency ones, are not always made explicit. One argument is that dependencies often yield unique solutions, in the form of a single dependency graph, where alternative CFG theories license multiple analyses. For example, for the VP ‘ate dinner at the table with a fork’ has a single dependency representation, but three different CFG parse trees: (1) a ‘flat’ representation where the root VP dominates (in a 4-ary branching treelet) the verb ‘ate’, the NP ‘dinner’, and the PPs ‘at the table’ and ‘with a fork’; (2) a hierarchical Chomskyan adjunction analysis with only binary branching structure; (3) a mixed two-level Penn Treebank-style representation, as in (1), except for ‘ate dinner’, which is represented as a VP

<sup>18</sup> See Kokkonidis (2008) or Andrews (2010) for some more details. We will just briefly mention two issues, one with Glue semantics generally and another with its application to dependency grammar in particular. The first concerns the resource logical aspects of the framework. Taken lexically, this means that “words contribute to the entire meaning exactly once. It is not possible to ignore them or to use their meaning more than once” (Müller, 2018: 226). There are a number of linguistic cases which seem to militate against this possibility. Expletives in non-pro drop languages, like English, do not contribute to the meaning of sentences except to fulfill a syntactic requirement on subject positions. Copredication, on the other hand, has been the utilised by Chomsky and others to argue that a word can contribute to the meaning of a sentence multiple times (Chomsky, 2000a; Collins, 2017; Vicente, 2021). Of course, copredication is a difficult phenomenon and poses a challenge for truth-conditional approaches as well.<sup>19</sup> Glue semantics might be versatile enough to be able to accommodate it, especially given its strong commitment to syntactic autonomy.<sup>20</sup> A slightly more pertinent concern is with the commutativity of linear logic. Since the semantic composition imposes no order constraints, this task is left to the syntax. In LFG, the c-structure contains linear order information (see Sect. 6.4). However, in dependency grammar, linearization needs a separate account. Thus, a commutative semantics paired with a commutative syntax potentially leaves an explanatory gap.



([V NP]), as per the Chomskyan analysis. In the context of many NLP applications, there may be no appreciable or useful differences between these analyses, and what they share may be captured sufficiently well by the single dependency graph.

In the NLP literature, two classes of dependency parsers are often discussed and contrasted: *graph-based parsers*, which search for an optimal parse by maximizing a chosen scoring function over the space of valid dependency graphs of the input; and *transition-based parsers*, which generate a parse by applying a predefined set of actions from an initial configuration, typically containing the root node. Tractability results seem to favor transition-based parsers overall (Eisenstein, 2019), but a more cogent argument for transition-based systems in this context may be their relative cognitive plausibility. Graph-based algorithms require that all words and all potential arcs in a sentence are held in memory and are available simultaneously for scoring. Instead, transition-based algorithms parse sentences incrementally, applying actions (e.g., create a dependency arc to the left versus right of the current word, starting at the root) that update a dependency representation, using a memory buffer and a stack. Shifting items from the buffer to the top of the stack and removing elements from the stack are two actions that the parser can perform: i.e., SHIFT and REDUCE, besides ARC-LEFT or ARC-RIGHT. The question arises whether these actions, as such, are a cognitively plausible way of implementing incremental dependency parsing (more on this below). But what is clear is that transition-based parsers scale up nicely to complex sentences, since the number of actions required to parse a sentence grows linearly as a function of input length. These properties of transition-based parsers make them convenient modeling tools, both in NLP, where efficiency is desirable, and in psycholinguistics, where item-by-item incrementality is a basic requirement (Covington, 2001). Our assumption is not that NLP models must be psychologically plausible, but that algorithms that are psychologically plausible may yield comparative efficiency gains, given the apparent efficiency of human parsing.

## 6 Psychological and Neural Plausibility

In Sects. 4 and 5, we have presented and discussed some of the arguments deployed to justify dependencies on empirical, theoretical, or computational grounds. Overall, on the basis of these arguments alone, a conclusive case for dependencies, against or over constituency structure, cannot be made. A critical step is missing: a discussion of the cognitive and neural plausibility of dependency grammar, also in comparison to constituency formalisms. To anticipate the argument in this section: there is some evidence for dependency structure in human sentence processing, which however equally supports constituency structure. According to the received position, this can be explained in terms of the notion of notational variance. We take a different route here. We argue that this situation presents the puzzle of how these two formalisms could or should be combined to account for syntactic processing data and raises a more fundamental issue with the nature of human syntactic competence.

## 6.1 Arguments and Evidence for Syntactic Dependencies

In the psycholinguistic literature, four main types of arguments have been developed in support of dependencies. One starts from the general observation that the human mind can represent hierarchical part-whole relations, as required by PSGs, and sets, as required by minimalist accounts. Moreover, various structural dependencies, such as among the nodes of single-rooted directed acyclic graphs, are cognitively possible, and are therefore ‘available’ for syntax (Hudson, 2016). These arguments would then proceed to take the additional step from cognitive possibility to cognitive plausibility or reality. Plausibility arguments build on various considerations, such as *parsimony* (e.g., that PSGs involve word class representations and non-terminal nodes, whereas dependency formalisms do not) and *learnability* (unsupervised inductive learning of syntactic dependencies from data is possible, Klein & Manning, 2004; it is not clear, however, whether those algorithms make realistic assumptions about human language learning, Clark, 2017). But parallel considerations could be developed for constituents and other grammatical structures (e.g., constructions), which appears to undermine the logic of most non-comparative plausibility arguments for dependencies. Another line of argument reconsiders traditional constituency tests to show that PSGs assume more structure than can be justified based on the tests alone, in particular subphrasal structure (Osborne, 2018). However, in our view, none of these arguments succeeds in establishing that dependency structures have greater a priori cognitive plausibility than constituency (note the comparative construction), let alone greater psychological reality. Cross-linguistic, psycholinguistic, and neurolinguistic data seem necessary to take this further step, or at least to more fully assess its feasibility.

The second type of argument is provided by growing cross-linguistic evidence in favor of dependency length minimization (DLM), which was mentioned in Sect. 4.1.1 (Ferrer-i-Cancho, 2004; Gildea & Temperley, 2007, 2010; Liu, 2008; Temperley, 2008; Futrell et al., 2015; 2015b; Ferrer-i-Cancho, 2016a; Liu et al., 2017; Gibson et al., 2019; Ferrer-i-Cancho et al., 2022). Here, the link with cognition lies in the fact that several models of human sentence processing predict that syntactic dependencies, in both production and comprehension, are harder to generate or process the longer they are, that is, the greater the number of words crossed over by dependency arcs. The observation that languages minimize dependency lengths would suggest that such parsing constraints shape languages, and that dependencies are therefore cognitively real. But a different reading of the data is possible, starting from the assumption that minimized dependency lengths also imply minimized LDDs and favor local constituent structure. Dependency locality is the notion that the linear distance between words linked in dependencies (dependency length) should be as short as possible. However, as noted in Sect. 4.1.1, the absolute minimum length of 1 corresponds to the principle of linear adjacency underlying constituency structure, leaving aside movement, LDDs, and related phenomena. Given minimized dependencies, it is still possible that the human mind/brain can generate and parse hierarchical constituent structures and that DLM facilitates constituency-based syntax and parsing, or that it enables them, i.e., that it is a precondition on the emergence of constituent structure.

The third argument is based on syntactic judgments. Hays (1964) already observed that “data about the structural intuitions of native speakers” are required to show that dependencies are cognitively viable: “dependency theory would be supported if many or most native speakers could agree on the central (governing) element in each utterance presented to them, and on the connections binding elements together. Phrase structure theory would be supported if they agreed on containments, e.g., that object-verb relations are closer than subject-verb relations, so that predicates are contained in sentences” (525). Hays voiced some skepticism on the possibility of collecting well-controlled data sets and especially on using them to decide between theories. However, Levelt (1974) laid out axiomatic theories for deriving predictions for syntactic cohesion judgments from PSGs (with or without transformations) and from dependency grammars. Intuitions about syntactic cohesion (namely, whether or not words or phrases belong together in a sentence) may be elicited and studied in different ways, all covered by Levelt’s theory.<sup>21</sup> Levelt intended this as an approach for adjudicating between alternative syntactic representations within a particular theory, or between alternative theories, e.g., PSGs and dependency grammars. He applied measurement theory to formally connect theoretical analyses of sentences to matrices of coherence values. For example, the analysis derived from a PSG entails that, for the sentence (Chomsky’s example) ‘John decided on the boat on the train’, ‘decided’ and ‘on the train’ have less cohesion than ‘decided’ and ‘on the boat’. This cohesion pattern is attributed to the relationship of constituency/inclusion, where ‘decide on the boat’ is a VP, to which the PP ‘on the train’ is attached as an adjunct. Levelt then uses experimental data to show that PSGs predict cohesion judgments poorly, especially as the syntactic hierarchy grows or becomes more complex, also in transformational versions of PSG. Instead, dependency grammars provide a better fit with cohesion data. To our knowledge, Levelt’s results have not been replicated, nor have there been (failed) replication attempts. But it is surprising that his formal, deductive method for testing the effects of alternative syntactic representations on cohesion data has not been more widely adopted in psycholinguistics. To date, this is arguably the clearest demonstration of the empirical superiority of dependency grammar over traditional PSGs vis-à-vis syntactic judgments.

The fourth argument is based on on-line behavioral evidence that human language processing tracks syntactic dependencies between words. For example, in a recent eye tracking study, Lopopolo et al. (2019) showed that, for each word in a stimulus sentence, the number of backward saccades made by readers at that word can be predicted by the number of left-hand-side dependents of that word. The effects of dependencies on regressions from a word were observed only for the preceding 4 words. These data seem consistent with the view that readers represent and track syntactic dependencies on-line, but they cannot exclude that constituent structure plays a role in language processing along with dependency relations (more on this below). Another possible source of relevant data is structural priming. It has been suggested that structural priming effects, where the processing of two subsequent sentences is affected by their structural/syntactic similarity (Bock, 1986; Brennan & Pykkänen,

<sup>21</sup> Levelt (2020) describes this theory as “the best-kept secret of the book” (Levelt, 1974), and adds that it is “as valid now as it was 45 years ago”.

2017), supports dependency structure (Hudson, 2017), especially for some grammatical relations such as subject-verb agreement (Haskell et al., 2010). There are equivalent ways of defining structures in constituency and dependency formalisms, explaining structural priming effects equally well. We are not aware of any structural priming studies testing diverging predictions from those formalisms.

## 6.2 Arguments and Evidence for Constituents and Hierarchies

In linguistic (meta-)theory, constituency-based and dependency-based formalisms are often considered as notational variants: the assumption is that there is no real dispute as to which one provides a (more) faithful representation of the syntactic structures of language. If they were indeed equivalent, ‘hybrid approaches’ combining them would be trivial. Yet, research in computational linguistics and NLP, in spite of widespread use of dependency parsers, testifies to the contrary: there is room and independent motivation for hybrid constituency-dependency parsers (e.g., see Hall et al., 2007), and hybrid approaches may even be a forced choice for modeling human syntactic competence and processing, if one accepts jointly all the arguments and all evidence currently available: syntactic dependencies (or equivalent functional-tier structures) could well be cognitively real, but so may be hierarchies and constituents. Research in theoretical linguistics, for example in LFG and the Parallel Architecture (Culicover & Jackendoff, 2005), points to a similar conclusion.

Frank et al. (2012) discuss computational and experimental studies indicating that hierarchical syntactic structure, as postulated by constituency-based grammars, may not necessarily be represented during on-line language processing, and that, instead, “sequential structure” (flat representations of grammatical or semantic dependencies between words) is often enough to explain human or model performance. Frank et al.’s discussion is sufficiently nuanced to be broadly consistent with hybrid accounts: they cannot exclude that hierarchical structure plays some role in language comprehension and production, such that, for example, any “evidence for hierarchical operations will only be found when the language user is particularly attentive, when it is important for the task at hand (e.g., in meta-linguistic tasks), and when there is little relevant information from extra-sentential/linguistic context” (4528). We are not aware of any current hybrid parsing models that make empirical predictions of this kind, testable using human data. But indeed, what should be tested experimentally is not whether dependencies and syntactic constituents co-exist—they most likely do—, but rather under what conditions the human language system commits to one set of algorithms and representations, in what conditions it switches to the other, and how. The results of experiments in cognitive neuroscience may shed some light on this question.

Pallier et al. (2011) was among the first studies to show that the BOLD fMRI signal increases with the size of constituents in sentences of fixed length. There were two distinct cortical networks showing this pattern: a fronto-temporal system, where the effect of constituent size was also produced by well-formed jaberwocky sentences; and a temporo-parietal network, where the effect was found for sentences containing real words. Participants performed sentence detection and word memory tasks. The tasks did not require them to understand the sentences or to parse their constituent

structures. Yet, constituent effects were found. Frank et al. (2012) view these results as evidence that “sequentially structured [not hierarchically structured] constituents are extracted even when this is not relevant to the task at hand”. Theoretically, it is not quite clear how one can have constituents, on the technical definition that Pallier and colleagues also adopt, without hierarchies. In any event, this study shows rather convincingly that the brain tracks constituent structure in real time, in the absence of specific task demands that would induce such parsing strategy (see also Chang et al., 2020 for recent converging results using the same paradigm).

Other experiments have found brain responses compatible with on-line tracking of constituent structures. Ding et al. (2016) provide MEG data indicating that cortical responses at different timescales track constituents at different hierarchical levels: i.e., words, phrases, and sentences. In 4-word sentences, where each word is presented at a regular rate (at 4 Hz, for monosyllabic words) and words form phrases in pairs (e.g., ‘Dry fur rubs skin’), MEG responses showed peaks precisely at 1 Hz (sentence), at 2 Hz (phrase, ‘dry fur’ or ‘rubs skin’), and at 4 Hz (word; for EEG data, see Ding et al., 2017; for an explicit computational model, reproducing the response patterns in MEG data, see Martin & Doumas, 2017). This study has been criticized on grounds that the same power spectra could also be reproduced by a neural network that represents word vectors sequentially, and so it does not explicitly encode hierarchical structure (Frank & Yang, 2018). More recent studies have tried to disentangle the hierarchical and the lexical/sequential accounts, finding that spectral responses are not accounted for by syntactic constituent tracking (Kalenkovich et al., 2022; Glushko et al., 2020). It remains unclear whether aspects of brain oscillations directly encode constituent structure (Tavano et al., 2021). Studies looking into other electrophysiological signals have reported responses consistent with hierarchical representations of constituents (e.g., see Nelson et al., 2017; for a commentary, see Chen et al., 2018) or have provided more direct and comparative evidence for it (Brennan et al., 2016). On the one hand, these results do not conclusively establish that constituent structure is routinely and automatically represented during on-line sentence processing. But on the other hand, they agree well with armchair arguments one could make about the human *capacity* to represent hierarchical constituent structure: it is not controversial that such a capacity exists and can be deployed on-line during sentence processing; what is at stake is precisely under what conditions that capacity is exercised and what (other) computational resources, beyond dependencies, are exploited when phrase structure is not built. This takes us to the closing theme: the architecture of syntax.

### 6.3 The Architecture of Syntax: An Integrative Hypothesis

The arguments and evidence reviewed above may be taken to suggest that neither constituency-based nor dependency-based frameworks alone are sufficient to model human syntactic competence and to explain all cross-linguistic, psycholinguistic, and neurolinguistic observations: they both seem necessary (and may be jointly sufficient) to achieve full explanatory adequacy. Moreover, the human brain seems capable of representing both constituent structure and dependency structure, as empirical data as well as linguistic practice indicate. One reasonable conclusion, then, is that human

syntactic competence in fact encompasses core properties of both constituency and dependency grammars, which may be exploited jointly or alternatively depending on the specific language Evans and Levinson (2009), on specific constructions in a given language, on the task at hand, or on other (yet unknown) conditions on representation and processing. A recent fMRI study supports the notion that dependency structure and phrase structure are derived on-line and represented in different cortical regions: the left anterior temporal pole and the left inferior frontal gyrus track dependency relations, whereas the left posterior superior temporal gyrus is sensitive to phrase structure (Lopopolo et al., 2020). It remains unclear what role these representations play in the architecture of syntax, especially in relation to processing: are dependency structure and phrase structure always derived together, sequentially or in parallel, or are they computed alternatively or jointly depending on yet unknown factors? This question has been overlooked in language science. We cannot offer a definite answer here, but we provide preliminary reasons why a productive research program could be developed on this basis.

The idea of hybrid constituency-dependency grammars and parsers is not new. Hudson (2010) among others has noted that there are no technical obstacles to adding dependency edges to the terminal nodes of a phrase structure tree. In the resulting hybrid structures, dependency and constituency relations are not redundant: this is a consequence of rejecting the received position, especially if one allows non-projective dependencies. In fact, it is easy to show that several linguistic frameworks already combine key elements of dependencies and phrase structure: e.g., X-bar theory, LFG ('c-structure' and 'f-structure'), dependency-constituency structure in dependency categorial grammar (Pickering & Barry, 1993), and phrase structure versus grammatical functions in *Simpler Syntax* (Culicover & Jackendoff, 2005). Klein and Manning's (2004) seminal model used constituent structure and the probability of head dependency relations to identify syntactic structure in complex sentences: they show that their dependency-constituency model outperforms dependency and constituency systems considered separately (for a discussion, see Lappin & Shieber, 2007). Hall et al. (2007) designed a parser that produces dependency and constituency structures, separately or in combination in a single process. The parsing accuracy performance was only slightly lower (1% or less) than that of non-hybrid representations in either format (for explorations of techniques for increasing accuracy, see Green & Zabokrtský, 2012).

The question is how hybrid grammars and parsers may be used to construct or constrain plausible models for human syntactic competence and processing. A viable architecture might be one in which, as in the Hall et al. (2007) model, constituency and dependency representations can be generated jointly. But as a default, the system will try to compute sentence meaning minimizing the amount of syntactic structure that is overtly represented (be it phrase structure or dependency structure), so long as an interpretation is assigned to the input. In this process, syntactic dependencies, in particular grammatical role labels (subject, object etc.), may be used to check on the compatibility of the interpretation pursued by the semantic system with bottom-up grammatical input cues (for a computational model of experimental data, see Michalon & Baggio, 2019; for background, reviews of evidence, and theoretical motivation, see Baggio, 2018). In general, interpretations can be computed without explicitly deriving

constituent structure, and without composing constituent meanings based on that, as compositionality would mandate. Dependency structure provides minimal structural and grammatical constraints in exactly those cases.

(5) The first case of AIDS was reported in 1975. (Pelletier, 1994)

We read (5) correctly as saying that the first AIDS case report was in 1975, and not that the first case of AIDS ever was reported only in 1975. It is not obvious that (and how) a compositional analysis, based on phrase structure, could license that reading: on such an analysis, ‘reported’ cannot modify ‘case’ (the first reported case of AIDS) *and* be part of the predicate VP. Here, one could exploit an “exclusively syntactic and lexical route to sentence-level semantic content” (Borg, 2012), but that would give an incorrect, unintended meaning. So, whereas a dependency graph for (5) would be fully consistent with the intended interpretation, a phrase structure tree would not. In a range of cases, computing constituent structure in the service of compositional interpretation does return the correct, intended meaning. But that is not a universal recipe for deriving contextually plausible interpretations. When would constituency parsing be actually performed? A shift to overt phrase structure representations and compositional interpretations may not be primarily driven by task demands, as the neuroscientific evidence reviewed above indicates. Our best guess is that properties of the input are the main factor. For example, the greater the structural complexity of a phrase or sentence (signaled by length in words and by the number of function words or morphemes), the greater the engagement of constituency representations and compositional processing on-line (Prinz, 2012; Baggio et al., 2012; Baggio, 2021).

This picture may turn out to be simplistic, or incorrect in other ways, but we put it forward as one example of the kind of integrative approach that would be required to explain the extant data.<sup>22</sup> We also note how this differs from a pluralistic stance: we are not merely arguing that neither dependencies nor phrase structure are sufficient, and that both are necessary; our point is that they ought to be effectively integrated into an architecture that explains exactly in what conditions the two representations are generated jointly or independently.

#### 6.4 The LFG in the Room

Before we conclude, it behooves us to address the connection(s) between what we have been advocating, namely the ‘integrative hypothesis’, and the theory which has been developed for decades under the banner of Lexical Functional Grammar (LFG).<sup>23</sup> There are many similarities between these projects. LFG was initially proposed as a means of capturing aspects of a broader class of language families and a wider range

<sup>22</sup> For instance, one avenue of argument we have not considered is whether dependency and phrase structure are equally supported statistically. There is some work to suggest that dependency structure is recoverable from corpora, see Ferrer-i-Cancho and Reina (2002) and Futrell et al. (2019) for relevant studies. See Eisner (2002) for a Bayesian statistical approach to lexicalized transformational grammar with phrasal categories and Thompson and Newport (2007) for a study extending work on transitional probabilities in word-segmentation to language learning of phrase structure *via* such probabilities. We thank an anonymous referee for directing us toward research on grammar induction.

<sup>23</sup> We are grateful to an anonymous referee for pressing us to explore this comparison.

of morphosyntactic phenomena, such as free word order languages and agglutination (Hale, 1983; Austin & Bresnan, 1996). Additionally, LFG was established on a platform of psycholinguistic plausibility (Bresnan, 1982) and is considered a ‘deep approach’ to NLP, since it aims to provide a tractable computational framework which is inspired by theoretical linguistics (Forst, 2011). However, it is generally not pitched at the level of individual theory, similarly to the Minimalist Program.

[T]he formal model of LFG is not a syntactic theory in the linguistic sense. Rather, it is an architecture for syntactic theory. Within this architecture, there is a wide range of possible syntactic theories and subtheories, some of which closely resemble syntactic theories within alternative architectures, and others of which differ radically from familiar approaches (Bresnan et al. 2016: 39)

This characterization opens up the possibility that our integrative hypothesis simply amounts to a version or particular instantiation of LFG.<sup>24</sup> We must urge against this interpretation: although there are a number of interesting connections, we do depart from the architecture of LFG in various respects.

The most profound connection might be LFG’s take on the notational variance of phrase structure and functional structure. LFG maintains that syntax comprises multiple interconnected levels of linguistic information. The most prominently articulated levels have been f-structure (functional), c-structure (constituent and category), and a-structure (argument). But LFG also contains p-structure (phonology and prosody), i-structure (information), s-structure (syntax-semantics interface), and m-structure (morphology) (Börjars, 2020). We will focus on f-structure and c-structure here since they map well onto the relationship between dependency and phrase structure.

F-structure represents the grammatical functions, such as subject and object, and is supposed to be invariant between languages. The standard means of representing f-structure is as an attribute-value matrix (AVM) or an unordered set of feature-value pairs, e.g., TENSE=FUTURE etc. Lexical items, such as nouns and verbs, also have a PRED feature with a unique semantic value. The f-structure shares some similarities with theta grids in Government and Binding theory and feature structures in unification-based accounts like Head-driven Phrase Structure Grammar. F-structure comes with a few conditions such as the ‘uniqueness condition’, or the requirement that every feature has exactly one value, and the ‘coherence condition’, which mandates that all argument functions occur in the value of a local PRED feature and that all functions with a PRED feature also have a  $\theta$ -role (thematic role).

C-structure is familiar from phrase structure grammar and is, in fact, a modified version of X-bar theory. However, as Börjars (2020) notes: “in comparison to other frameworks, LFG’s approach to X-bar syntax is unorthodox in that, for instance, nonbinary branching as well as exocentric categories is permitted” (157) and “that all nodes, including preterminal and head nodes, are optional” (159). Unlike f-structure, c-structure can and does vary cross-linguistically. One of the guiding motivations for LFG was to account for nonconfigurational languages like Warlpiri in which word

<sup>24</sup> More precisely, the overarching architecture is a ‘correspondence architecture’ of which LFG is one member. Our account might indeed fall under this general approach.



order is much more flexible and the phrase structure appears to be more ‘flat’, as in dependency analysis. Multiple c-structures can map onto the same f-structure.

Essentially, LFG rejects the idea that the phrase structure and dependency graph are notational variants. In fact, it motivates their union via interface principles. Much like dependency grammar, LFG is a lexicalized formalism, because “every partial tree licensed by a grammar rule contains a lexical element” (Müller, 2018: 392). Further similarities include relationships between projects like the Pargram for LFG<sup>25</sup> and the Universal Dependencies for dependency grammar, both of which aim to find common or universal syntactic structures (and labelling conventions) across languages by using their respective formalisms as a guide. Haug (2012) has even attempted to convert a dependency treebank into an LFG treebank.

There are, however, important considerations that prevent a complete reduction of our integrative hypothesis and LFG. In LFG’s architecture, c-structure interfaces with the phonological component of the grammar, whereas f-structure is crucial for semantic interpretation. Indeed, as mentioned in Sect. 5.2, glue semantics has been proposed as a semantic formalism for both f-structure and dependency grammar. Nevertheless, we are not committed to the requirement that semantic interpretation interacts *only* with the dependency aspects of syntax. On our proposal (see 6.3), the phrase structure dovetails more clearly with compositional semantics than does the dependency grammar. For this reason, it must retain its relationship with semantics, as in frameworks like GB, where the position of arguments in a tree play a prominent role in semantic composition. Thus, our framework allows for a degree of redundancy in syntactic representation (and semantic overlay) between these components, while LFG has a more clear division of labour: the c-structure captures linear precedence, dominance relations, such as c-command, and constituent structure; the f-structure represents grammatical functions and features such as number and case.<sup>26</sup>

Naturally, f-structures and dependency graphs are related in the information that they contain respectively. As Müller (2018: 366) writes, “[a]n unordered dependency graph assigns grammatical functions to a dependent of a head and hence it is similar in many respects to an LFG f-structure”, but then he goes on to note that there are certain elements of dependency graphs not present in f-structures: non-predicative prepositions are one example. Moreover, dependency graphs can encode argument structure usually reserved for a-structures in LFG. On the other hand, f-structures generally contain much more information than dependency graphs. In addition, UD favours content words over function words as heads, which means that “unlike in LFG representations, prepositional phrases are headed by nouns, numeral phrases are headed by nouns, and auxiliaries and copulas are always dependents, rather than heads” (Przepiórkowski & Patejuk 2020: 195). Interestingly, in providing an algorithm for deriving dependency structures from LFG analyses, Meurer (2017) has argued that c-structures are a better basis for the conversion than f-structures:

At a first glance, it is the f-structures that resemble dependency structures most (...). This correspondence is however not perfect; f-structure PRED values cannot

<sup>25</sup> <https://pargram.w.uib.no/>, <https://wiki.uni-konstanz.de/pargram/>.

<sup>26</sup> Although LFG has flexibility in this regard. See Asudeh and Toivonen (2014) for a focused discussion of this possibility.

easily be related to surface words (which the dependency nodes should consist of), because the projection relation is not injective (185).

Meurer therefore starts off with c-structure and projection in his algorithm, resulting in projective dependency structures. But he also demonstrates that the latter can, in turn, be converted into non-projective dependencies and UD-style ones. This is all to say that the relationships between our view and LFG, between dependency graphs and f-structures, between UD and Pargram, are intricate and complex. There are no straightforward mappings, although similarities abound. LFG definitely encompasses the same spirit as our proposal and there is much to learn from its well-established results.

## 7 Conclusion

In this article, we have made a case for the cognitive significance, divergence, and integration of dependency and constituency structures *contra* the received position on their notational variance. In this task, we have reviewed literature from formal language theory, empirical linguistics, and cognitive neuroscience. There remains a lot of interdisciplinary work to be done on this issue, but we hope to have reignited the conversation by assessing multiple arguments and evidences from a range of theoretical perspectives.

**Funding** Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital)

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Andrews, A. (2010). Propositional glue and the projection architecture of LFG. *Linguistics and Philosophy*, 33, 141–170.
- Asher, N. (2011). *Lexical meaning in context: A web of words*. Cambridge University Press.
- Austin, P., & Bresnan, J. (1996). Non-configurationality in Australian aboriginal languages. *Natural Language and Linguistic Theory*, 14, 215–268.
- Asudeh, A. (2004). Resumption as resource management. Ph.d. thesis, Stanford University.
- Asudeh, A. (2012). *The logic of pronominal resumption*. Oxford University Press.
- Asudeh, A. (2022). Glue semantics. *Annual Review of Linguistics*, 8, 321–341.
- Asudeh, A., & Toivonen, I. (2014). With lexical integrity. *Theoretical Linguistics*, 40, 175–186.
- Baggio, G. (2018). *Meaning in the Brain*. MIT Press.
- Baggio, G. (2021). Compositionality in a parallel architecture for language processing. *Cognitive Science*, 45(5), e12949.

- Baggio, G., Van Lambalgen, M., & Hagoort, P. (2012). The processing consequences of compositionality. In M. E. Werning, W. E. Hinzen, & E. E. Machery (Eds.), *The Oxford handbook of compositionality* (pp. 655–672). Oxford University Press.
- Berwick, R. (1984). Strong generative capacity, weak generative capacity, and modern linguistic theories. *Computational Linguistics*, 10(3–4), 189–202.
- Bickerton, D. (1981). *Roots of language*. Karoma Publishers, Inc.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18(3), 355–387.
- Borg, E. (2012). *Pursuing meaning*. Oxford University Press.
- Börjars, K. (2020). Lexical-functional grammar: An overview. *Annual Review of Linguistics*, 6(1), 155–172.
- Bouma, G., van Noord, G., & Malouf, R. (2000). Alpino: Wide-coverage computational analysis of Dutch. In *Computational linguistics in the Netherlands 2000: selected papers from the 11th CLIN Meeting* (pp. 45–59).
- Brennan, J. R., Stabler, E. P., Van Wagenen, S. E., Luh, W. M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, 157, 81–94.
- Brennan, J. R., & Hale, J. T. (2019). Hierarchical structure guides rapid linguistic predictions during naturalistic listening. *PLoS ONE*, 14(1), e0207741.
- Brennan, J. R., & Pyllkänen, L. (2017). MEG evidence for incremental sentence composition in the anterior temporal lobe. *Cognitive Science*, 41, 1515–1531.
- Bresnan, J. (Ed.). (1982). *The mental representation of grammatical relations*. MIT Press.
- Bresnan, J., Asudeh, A., Toivonen, I., & Wechsler, S. (2016). *Lexical-functional syntax*. Wiley.
- Bresnan, J., Kaplan, R., Peters, S., & Zaenen, A. (1982). Cross-serial dependencies in Dutch. In *The formal complexity of natural language* (pp. 286–319.) Springer.
- Bröker, N. (2003). Formal foundations of dependency grammar. In Ágel, V., Eichinger, L. M., Eroms, H.-W., Hellwig, P., Jürgen H. H. & Lobin, H. (Eds.), *Dependenz und Valenz/Dependency and valency: Ein internationales Handbuch der zeitgenössischen Forschung / An international handbook of contemporary research*, vol. 1 (Handbücher zur Sprach- und Kommunikationswissenschaft 25) (pp. 294–310). Walter de Gruyter.
- Chang, C. H., Dehaene, S., Wu, D. H., Kuo, W. J., & Pallier, C. (2020). Cortical encoding of linguistic constituent with and without morphosyntactic cues. *Cortex*.
- Chen, X., & Gerdes, K. (2017). Classifying languages by dependency structure: Typologies of delexicalized Universal Dependency treebanks. In *Proceedings of the 4th international conference on dependency linguistics*, (pp. 54–63). Assoc. Comput. Linguist.
- Chen, X., Gómez-Rodríguez, C., & Ferrer-i-Cancho, R. (2018). A dependency look at the reality of constituency. *Glottometrics*, 40, 104–106.
- Chomsky, N. (1963). Formal properties of grammars. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology II* (pp. 323–418). Wiley.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.
- Chomsky, N. (1972). *Studies on semantics in generative grammar*. Mouton.
- Chomsky, N. (1981). *Lectures on government and binding*. Foris.
- Chomsky, N. (1995). *The minimalist program*. MIT Press.
- Chomsky, N. (2000). *New horizons in the study of mind and language*. Cambridge University Press.
- Chomsky, N. (2000). Minimalist inquiries. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (pp. 89–155). MIT Press.
- Clark, A. (2017). Computational learning of syntax. *Annual Review of Linguistics*, 3, 107–123.
- Collins, J. (2017). The copredication argument. *Inquiry*, 60(7), 675–702.
- Covington, M. A. (2001). A fundamental algorithm for dependency parsing. In Miller, J. A., & Smith, J. W. (Eds.), *Proceedings of the 39th annual ACM Southeast conference* (Vol. 1, pp. 95–102).
- Culy, C. (1985). The complexity of the vocabulary of Bambara. *Linguistics and Philosophy*, 8, 345–351.
- Culicover, P., & Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.
- Dalrymple, M., Lowe, J., and Mycock, L. (2019). The Oxford Reference Guide to Lexical Functional Grammar (Oxford, 2019; online edn, Oxford Academic, 19 Dec. 2019), <https://doi.org/10.1093/oso/9780198733300.001.0001>
- de Marneffe, M. C., & Nivre, J. (2019). Dependency grammar. *Annual Review of Linguistics*, 5, 197–218.
- Debusmann, R. (2000). An introduction to dependency grammar. *Hausarbeit für das Hauptseminar Dependenzgrammatik SoSe 99* (pp. 1–16). Universität des Saarlandes.

- Debusmann, R., & Kuhlmann, M. (2010). Dependency grammar: Classification and exploration. In *Resource-adaptive cognitive processes* (pp. 365–388). Springer.
- Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in Human Neuroscience, 11*, 481.
- Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*(1), 158.
- Duchier, D. (2000). Lexicalized syntax and topology for non-projective dependency grammar. In *Joint conference on formal grammars and mathematics of language*, FGMOL'01, Helsinki.
- Eisenstein, J. (2019). *Introduction to natural language processing*. The MIT Press.
- Eisner, J. (2002). Discovering syntactic deep structure via Bayesian statistics. *Cognitive Science, 26*, 268–2555.
- Evans, N., Levinson, S.C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32 (5), 429–492.
- Ferrer-i-Cancho, R. (2004). Euclidean distance between syntactically linked words. *Physical Review E, 70*, 056135.
- Ferrer-i-Cancho, R. (2006). Why do syntactic links not cross? *Europhysics Letters, 76*(6), 12–28.
- Ferrer-i-Cancho, R. (2016a). Non-crossing dependencies: Least effort, not grammar. In *Towards a theoretical framework for analyzing complex linguistic networks* (pp. 203–234). Springer.
- Ferrer-i-Cancho, R. (2016b). Kauffman's adjacent possible in word order evolution. In Roberts, S. G., Cuskley, C., McCrohon, L., Barceló-Coblijn, L., Fehér, O. & Verhoef, T. (Eds.), *The evolution of language: Proceedings of the 11th international conference* (EVOLANG11).
- Ferrer-i-Cancho, R., & Reina, F. (2002). Quantifying the semantic contribution of particles. *Journal of Quantitative Linguistics, 9*, 35–47.
- Ferrer-i-Cancho, R., Gómez-Rodríguez, C., Esteban, J., & Alemany-Puig, L. (2022). Optimality of syntactic dependency distances. *Physical Review E, 105*, 014308.
- Frank, S., Bod, R., & Christiansen, M. (2012). How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences, 279*, 4522–4531.
- Frank, S. L., & Yang, J. (2018). Lexical representation explains cortical entrainment during speech comprehension. *PloS ONE, 13*(5).
- Forst, M. (2011). Computational aspects of lexical functional grammar. *Language and Linguistics Compass, 5*(1), 1–18.
- Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences, USA, 112*(33), 10336–10341.
- Futrell, R., Mahowald, K., & Gibson, E. (2015b). Quantifying word order freedom in dependency corpora. In Hajicová, E., & Nivre, J., (Ed.), *Proceedings of the 3rd international conference on dependency linguistics* (pp. 91–100). Assoc. Comput. Linguist.
- Futrell, R., Qian, P., Gibson, E., Fedorenko, E., & Blank, I. (2019). Syntactic dependencies correspond to word pairs with high mutual information. In *Proceedings of the fifth international conference on dependency linguistics* (Depling, SyntaxFest 2019).
- Gaifman, H. (1961). Dependency systems and phrase structure systems. Technical report P-2315, The RAND Corporation.
- Garde, P. (1977). Ordre linéaire et dépendance syntaxique?: contribution á une typologie. *Bull. Soc. Ling. Paris, 72*(1), 1–26.
- Gell-Mann, M., & Ruhlen, M. (2011). The origin and evolution of word order. *Proceedings of the National Academy of Sciences of the USA, 108*(42), 17290–17295.
- Gibson, E., Futrell, R., Piandadosi, S. T., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences, 25*(5), 389–407.
- Gildea, D., & Temperley, D. (2007). Optimizing grammars for minimum dependency length. In *Proceedings of the 45th annual meeting of the association of computational linguistics* (pp. 184–191).
- Gildea, D., & Temperley, D. (2010). Do grammars minimize dependency length? *Cognitive Science, 34*(2), 286–310.
- Givón, T. (1975). Serial verbs and syntactic change: Niger-Congo. In C. Li (Ed.), *Word order and word order change* (pp. 47–112). University of Texas Press.
- Glushko, A., Poeppel, D., & Steinhauer, K. (2020). Overt and covert prosody are reflected in neurophysiological responses previously attributed to grammatical processing. bioRxiv.

- Gómez-Rodríguez, C. (2016). Restricted non-projectivity: Coverage vs. efficiency. *Computational Linguistics*, 42(4), 809–817.
- Gómez-Rodríguez, C., & Ferrer-i-Cancho, R. (2017). Scarcity of crossing dependencies: A direct outcome of a specific constraint? *Physical Review E*, 96(6), 062304.
- Gotham, M., & Haug, D. (2018). Glue semantics for universal dependencies. In Butt, M., & King, T. H. (Eds.), *Proceedings of the LFG18 conference* (pp. 208–226). CSLI Publications.
- Green, N. D., & Žabokrtský, Z. (2012). Hybrid combination of constituency and dependency trees into an ensemble dependency parser. In *Proceedings of the workshop on innovative hybrid approaches to the processing of textual data* (pp. 19–26). Association for Computational Linguistics.
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of language* (pp. 73–113). MIT Press.
- Hale, K. (1983). Warlpiri and the grammar of non-configurational languages. *Natural Language and Linguistic Theory*, 1, 5–47.
- Hall, J., Nivre, J., & Nilsson, J. (2007). A hybrid constituency-dependency parser for Swedish. In *Proceedings of the 16th nordic conference of computational linguistics (NODALIDA 2007)* (pp. 284–287).
- Haskell, T. R., Thornton, R., & MacDonald, M. C. (2010). Experience and grammatical agreement: Statistical learning shapes number agreement production. *Cognition*, 114(2), 151–64.
- Haug, D. (2012). From dependency structures to LFG representations. In Butt, M., & T. H. King (Eds.), *Proceedings of the LFG'12 conference* (pp. 271–291). Udayana University.
- Hays, D. (1960). Grouping and dependency theories. Research Memorandum RM-2538, The RAND Corporation.
- Hays, D. (1964). Dependency theory: A formalism and some observations. *Language*, 40, 511–525.
- Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford: Blackwell.
- Hudson, R. (1990). *English word grammar*. Basil Blackwell.
- Hudson, R. (2007). *Language networks: The new word grammar*. Oxford University Press.
- Hudson, R. (2010). *An introduction to word grammar*. Cambridge University Press.
- Hudson, R. (2016). Dependency grammar. In A. Hippisley & G. Stump (Eds.), *The Cambridge handbook of morphology* (pp. 657–682). Cambridge University Press.
- Hudson, R. (2017). Structural priming supports grammatical networks. *Behavioral and Brain Sciences*, 40, e295.
- Jackendoff, R. (1977). *X syntax: A study of phrase structure*. MIT Press.
- Jacobson, R. (1958/1984). Morphological observations on Slavic declension (the structure of Russian case forms). In Waugh, L. R., & Halle, M. (Eds.), *Roman Jakobson. Russian and Slavic grammar: Studies 1931–1981* (pp. 105–133). Mouton de Gruyter.
- Johnson, K. (2015). Notational variants and invariance in linguistics. *Mind and Language*, 30(2), 162–186.
- Jurafsky, D., & Martin, J. (2021). *Speech and language processing*. Prentice Hall.
- Kahane, S. (1996). If HPSG were a dependency grammar. In *Proceedings of the 3rd TALN conference* (pp. 45–49). Marseille.
- Kahane, S. (1997). Bubble trees and syntactic representations. In *Proceedings of the 5th meeting of mathematics of language* (pp. 70–76). Saarbrücken, Ger.: Dtsch. Forsh. Künstliche Intell.
- Kalenkovich, E., Shestakova, A., & Kazanina, N. (2022). Frequency tagging of syntactic structure or lexical properties; a registered MEG study. *Cortex*, 146, 24–38.
- Kamp, H. (1981). A theory of truth and semantic representation. In J. A. K. Groenendijk, T. M. V. Janssen, & M. B. J. Stokhof (Eds.), *Formal methods in the study of language: Mathematical centre tracts 135* (pp. 227–322). Mathematical Centre.
- Klein, D., & Manning, C. (2004). Corpus-based induction of syntactic structure: models of dependency and constituency. In *Proceedings of the 42nd annual conference of the association for computational linguistics*, art. 478. Stroudsburg.
- Kokkonidis, M. (2008). First-order glue. *Journal of Logic, Language and Information*, 17, 43–68.
- Kromann, M. (2003). The Danish dependency treebank and the DTAG treebank tool. In *Proceedings of the 2nd workshop on treebanks and linguistic theories*, (pp. 217–22). Väsjö University Press.
- Kübler, S., McDonald, R., & Nivre, J. (2009). *Dependency parsing*. Morgan & Claypool.
- Kuhlmann, M. (2013). Mildly non-projective dependency grammar. *Computational Linguistics*, 39(2), 355–387.
- Kuroda, S. (1976). A typological study of phrase structure languages. *Information and Control*, 30, 307–379.
- Lappin, S., & Shieber, S. M. (2007). Machine learning theory and practice as a source of insight into universal grammar. *Journal of Linguistics*, 43(2), 393–427.

- Levelt, W. J. (1974). *Formal grammars in linguistics and psycholinguistics: Vol. III*, Psycholinguistic applications: John Benjamins.
- Levelt, W. J. (2020). On becoming a physicist of mind. *Annual Review of Linguistics*, 6, 1–23.
- Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, 9(2), 159–191.
- Liu, H., Xu, C., & Liang, J. (2017). Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, 21, 171–193.
- Lopopolo, A., Frank, S. L., van den Bosch, A., & Willems, R. (2019). Dependency parsing with your eyes: Dependency structure predicts eye regressions during reading. In *Proceedings of the workshop on cognitive modeling and computational linguistics* (pp. 77–85).
- Lopopolo, A., van den Bosch, A., & Willems, R. M. (2020). Distinguishing syntactic operations in the brain: Dependency and phrase-structure parsing. *Neurobiology of Language*. <https://doi.org/10.1162/nol00029>
- Manaster-Ramer, A., & Kac, M. (1990). The concept of phrase structure. *Linguistics and Philosophy*, 13(3), 325–362.
- Martin, A., & Baggio, G. (2019). Modelling meaning composition from formalism to mechanism. *Philosophical Transactions of the Royal Society B*, 375, 20190298.
- Martin, A. E., & Dumas, L. A. (2017). A mechanism for the cortical computation of hierarchical linguistic structure. *PLoS Biology*, 15(3), e2000663.
- Mel'cuk, I. (1988). *Dependency syntax: Theory and practice*. State University Press of New York.
- Mel'cuk, I. (2011). Dependency in language. In *Proceedings of dependency linguistics 2011*, Barcelona.
- Meurer, P. (2017). From LFG structures to dependency relations. In Rosen, V., & Smedt, K. (Eds.), *Proceedings of the tenth international conference on language resources and evaluation, LREC 2016* (pp. 1659–1666). ELRA, European Language Resources Association.
- Michalon, O., & Baggio, G. (2019). Meaning-driven syntactic predictions in a parallel processing architecture: Theory and algorithmic modeling of ERP effects. *Neuropsychologia*, 131, 171–183.
- Miller, P. (1999). *Strong generative capacity*. CSLI Publications.
- Muller, S. (2018). *Grammatical theory: From transformational grammar to constraint-based approaches*. Berlin: Language Science Press.
- Nelson, M. J., El Karoui, I., Gibber, K., Yang, X., Cohen, L., Koopman, H., Cash, S. S., Naccache, L., Hale, J. T., Pallier, C., & Dehaene, S. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, 114(18), E3669–E3678.
- Nichols, J. (1986). Head-marking and dependent-marking grammar. *Language*, 62, 56–119.
- Nivre, J. (2008). Algorithms for deterministic incremental dependency parsing. *Computational Linguistics*, 34(4), 513–553.
- Oflazer, K., Say, B., Hakkani-Tur, D., & Tür, G. (2003). Building a Turkish treebank. In A. Abeille (Ed.), *Treebanks: Building and using parsed corpora* (pp. 261–77). Kluwer.
- o'Grady, W. (1998). The syntax of idioms. *Natural Language & Linguistic Theory*, 16(2), 279–312.
- Osborne, T. (2005). Beyond the constituent: A dependency grammar analysis of chains. *Folia Linguistica*, 39(3–4), 251–297.
- Osborne, T. (2014). Dependency grammar. In A. Carnie, Y. Sato, & D. Siddiqi (Eds.), *The Routledge handbook of syntax* (pp. 604–627). Routledge.
- Osborne, T. J. (2018). Tests for constituents: What they really reveal about the nature of syntactic structure. *Language Under Discussion*, 5(1), 1–41.
- Osborne, T., & Gerdes, K. (2019). The status of function words in dependency grammar: A critique of Universal Dependencies (UD). *Glossa: A Journal of General Linguistics*, 4(1), 17, 1–28.
- Osborne, T., Putnam, M., & Gross, T. M. (2011). Bare phrase structure, label-less trees, and specifier-less syntax. Is minimalism becoming a dependency grammar? *The Linguistic Review*, 28(3), 315–364.
- Osborne, T., Putnam, M., & Gross, T. (2012). Catenae: Introducing a novel unit of syntactic analysis. *Syntax*, 15(4), 354–396.
- Pallier, C., Devauchelle, A., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences of the United States of America*, 108(6), 2522–2527.
- Pelletier, J. (1994). The principle of semantic compositionality. *Topoi*, 13(1), 11–24.
- Penn, G. (2012). Computational linguistics. In R. Kempson, T. Fernando, & N. Asher (Eds.), *Philosophy of linguistics* (pp. 143–173). Elsevier B.V.

- Peters, S., & Ritchie, R. (1973). On the generative power of transformational grammars. *Information Sciences*, 6, 49–83.
- Pickering, M., & Barry, G. (1993). Dependency categorial grammar and coordination. *Linguistics*, 31(5), 855–902.
- Pitler, E., Kannan, S., & Marcus, M. (2013). Finding optimal 1-endpoint-crossing trees. *Transactions of the Association for Computational Linguistics*, 1, 13–24.
- Prinz, J. (2012). Regaining composure: A defense of prototype compositionality. In M. E. Werning, W. E. Hinzen, & E. E. Machery (Eds.), *The Oxford handbook of compositionality* (pp. 437–453). Oxford University Press.
- Przepiórkowski, A., & Patejuk, A. (2020). From lexical functional grammar to enhanced universal dependencies: The UD-LFG treebank of Polish. *Language Resources and Evaluation*, 54, 185–221.
- Rambow, O., & Joshi, A. (1997). A formal look at dependency grammars and phrase-structure grammars, with special consideration of word-order phenomena. *Recent Trends in Meaning-Text Theory*, 39, 167–190.
- Reinhart, T. (1976). The syntactic domain of anaphora. Doctoral dissertation, Massachusetts Institute of Technology.
- Robinson, J. (1970). Dependency structures and transformational rules. *Language*, 46, 259–285.
- Sgall, P., Hajicova, E., & Panevova, J. (1986). *The meaning of the sentence in its semantics and pragmatic aspects*. D. Reidel.
- Shieber, S. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, 8, 333–343.
- Stabler, E. (2019). Three mathematical foundations for syntax. *Annual Review of Linguistics*, 5(16), 18.
- Szabó, Z. (2000). *Problems of compositionality*. Routledge Press.
- Tavano, A., Blohm, S., Knoop, C.A., Muralikrishnan, R., Scharinger, M., Wagner, V., Thiele, D., Ghitza, O., Ding, N., Menninghaus, W. & Poeppel, D., (2021). Neural harmonics of syntactic structure. bioRxiv (pp. 2020-04).
- Temperley, D. (2008). Dependency-length minimization in natural and artificial languages. *Journal of Quantitative Linguistics*, 15(3), 256–282.
- Tesnieré, L. (1959). *Elements de Syntaxe Structurale*. Editions Klincksieck.
- Thompson, S., & Newport, E. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, 3(1), 1–42.
- van Rooij, I., & Baggio, G. (2021). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science*, 16(4), 682–697.
- Vicente, A. (2021). Chomskyan arguments against truth-conditional semantics based on variability and co-predication. *Erkenntnis*, 86, 919–940.
- Vijay-Shanker, K., & Weir, D. J. (1994). The equivalence of four extensions of context-free grammars. *Mathematical Systems Theory*, 27(6), 511–546.
- Zeevat, H. (1989). A compositional approach to discourse representation theory. *Linguistics and Philosophy*, 12(1), 95–131.
- Zeevat, H. (2006). Freezing and marking. *Linguistics*, 44(5), 1095–1111.