

# Mutual dependency and Word Grammar: headedness in the noun phrase

Nikolas Gisborne

Linguistics and English Language  
School of Philosophy, Psychology and Language Sciences  
University of Edinburgh  
n.gisborne@ed.ac.uk

## Abstract

Most Dependency Grammars reject mutual dependency but Word Grammar allows it, with Hudson (2004) exploiting it to analyse Determiner+Noun constructions. I discuss whether mutual dependency is desirable in Word Grammar and whether it is necessary in the treatment of Determiner+Noun structures. Many Dependency Grammar theories assume that the common noun is the head; early Word Grammar (Hudson, 1984) on the other hand argued for Determiner as head, and only more recently for mutual dependency. Mutual dependency is not permitted in most Dependency Grammars for formal reasons, because it violates the usual acyclicity constraint. However, natural language requires some relaxation of formal constraints on representations. I argue against mutual dependency in Word Grammar and also argue that it is not necessary in the analysis of Determiner+Noun; however, my arguments come from within Word Grammar and are based on its cognitive assumptions about how grammar is represented in the mind, rather than being based on formal criteria. My aim is to show that within a cognitive theory, constraints on representations can and should be stated in terms of the nature of the human cognitive system.

## 1 Introduction

Within a grammar that rejects exocentric analyses, there are three possible syntactic structures that can be assigned to the phrase *the dog*: (i) *the* depends on *dog*; (ii) *dog* depends on *the*; and (iii) the words are mutually dependent. The first two choices make no particular or unusual claims about the nature of Dependency Grammars (DGs): DGs are a class of grammar where there are pairwise relations between words. Both (i) and (ii) are compatible with constrained DGs and neither introduces formal violations of DG architecture. On the other hand, (iii) violates most dependency architectures because, apart from Word Grammar (WG), DGs adopt an acyclicity constraint.

Robinson (1970, 260) presents a series of axioms for a DG, given in (1). She states that the dependency relation is ‘transitive, irreflexive, and anti-symmetric’. She gives these as the ‘axioms of the theory which was advocated by Tesnière (1953), (1959) and formalized by Hays (1964) and Gaifman (1965).’

- (1) a. one and only one element is independent;
- b. all others depend directly on some element;
- c. no element depends directly on more than one other; and
- d. if *A* depends direct on *B* and some element *C* intervenes between them (in the linear order of the string), then *C* depends directly on *A* or *B* or some other intervening element.

Robinson’s axioms (1a-c) define a DG as a rooted tree, with (1d) also enforcing projectivity. Not every dependency theory subscribes to all of these axioms, but they offer a starting point. Ballesteros and Nivre (2013, 6) and McDonald and Nivre (2011, 202) describe similar constraints, with the latter telling us that

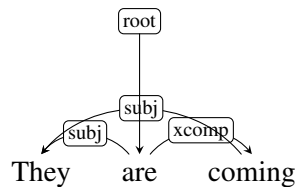


Figure 1: Analysis of *They are coming*.

the requirements in (1a-c) are ‘consistent with most formal theories’, such as Functional Generative Description (Sgall et al., 1986) and Meaning-Text Theory (Mel’čuk, 1988). The theories underwriting Osborne (2019) and Järvinen and Tapanainen (1998) also fit. McDonald and Nivre (2011) identify Word Grammar as a theory which is an exception in that it relaxes (1c), as does Anderson (2006)’s theory.

Figure 1 shows multiheadedness in a Word Grammar analysis of *they are coming*, using the analysis of Hudson (1984) and (1990). It represents *are* as the root of the sentence and *they* as the subject of both *are* and *coming*. Languages with agreement between syntactic subjects and predicative participles show why there has to be a syntagmatic relationship between *coming* and *they* in *they are coming*. In the absence of a syntactic relationship, what would, or could, carry the agreement in an example such as the French *elles-FEM.PL sont venues-FEM.PL*? As a result of this multiheadedness, if WG did not have mutual dependency, its representations would be directed acyclic graphs rather than dependency trees. However, as I have said, WG allows mutual dependency as well, including between Determiner (D) and Noun (N), making its representations general graphs. It is the point of this paper to show that this causes problems for WG theory, even given WG’s cognitive architecture, and that mutual dependency is not necessary in the case of D+N (Hudson, 2004).

Returning to the axioms in (1), it is worth briefly noting that many theories reject the projectivity constraint in (1d). For example, Mel’čuk (2014, 21) points out that the Latin sentence in (2) is non-projective due to the discontinuous relationship between *meas* and *nugas*.

- (2) Tu      *solebas*   **meas**            esse   aliquid            *putare*   **nugas**  
 you-NOM used-2SG my-FEM.PL.ACC be-INF something-NOM think-INF trifles-FEM.PL.ACC  
 ‘You used to think that my trifles were something’

The dependency between *solebas* and *putare* crosses the dependency between *meas* and *nugas* in a violation of projectivity. Many natural languages have such projectivity violations built in and one of the advantages of a non-projective DG is that it allows a direct, surface-level, representation of the syntactic structure of languages with discontinuous word order.

Constraints on WG representations are discussed in Hudson’s monographs. In Hudson (1984, 98-9) there is the Adjacency Principle, combined with the Priority to the Bottom Principle, which allows non-projectivity in extraposition; in Hudson (1990, 144ff.) there is a simplified version of the Adjacency Principle, but revised to allow multiple heads, retaining a version of Priority to the Bottom; in Hudson (2007) there is a theory of word order that dissociates dependencies from landmarks, with landmarks being responsible for word order. In Hudson (2010) this theory is developed and refined so that landmarks have to be projective, not dependencies themselves. Therefore, in WG projectivity is revised to allow extraction and, as we have seen, the constraint in (1c) is relaxed to permit structure sharing in predicative complementation, raising and control structures.<sup>1</sup>

What about mutual dependency, which Hudson (1990, 197) introduces in his analysis of relative clauses, and which is disallowed in other dependency theories? To the best of my knowledge, there are no mutual dependency analyses in Hudson (1984), or the papers that appeared between Hudson (1984) and Hudson (1990). I also have not found any particular arguments for mutual dependency: as far as I can see, it was adopted without much discussion, although the argument would be that looping

<sup>1</sup>Although I should note that Creider and Hudson (2006) introduce covert words into the WG ontology to handle a subtype of the infinitival construction in Ancient Greek, which introduces a further difference between WG and other DGs: as a consequence of this move, WG allows pairwise dependencies between realised and unrealised words.

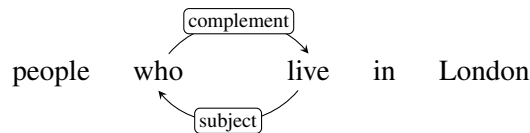


Figure 2: Mutual dependency in syntax

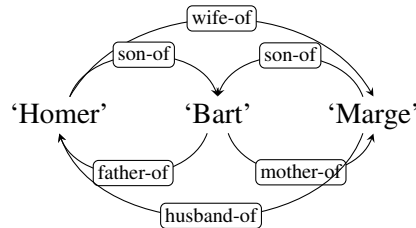


Figure 3: Marge and Homer: husband and wife;  
Bart has two parents

structures are inherent in networks, and in cognition. Hudson's representation for the phrase *people who live in London* (Hudson, 1990, 197) is given in Figure 2. Should structures like the one in Figure 2 be allowed in the syntax? This depends on the premises underlining the theory, and whether such structures are consistent with how the theory works.

WG is explicitly a cognitive theory (Hudson, 1984, 31-35) where language is analysed as part of a mental/cognitive network. Mutual dependency is a type of loop, and loops clearly exist in both network structures and mental networks. Hudson (2010, 49) makes this clear in his representations of family structure such as the partial structure of the Simpsons' family in Figure 3, where he shows the mutual relationship between Homer and Marge, and between each of them and Bart.<sup>2</sup> The point is that such structures are simply part of how a cognitive network must be structured.

The diagrams in Figures 2 and 3 are partially simplified, because in a full WG network they would also show classification relationships. The WG network is a classified network, with both nodes and arcs classified, where 'isa', the predicate of default inheritance, is a primitive. Given the network approach to cognition, the question for WG is whether mutual dependencies are cognitively plausible, not whether they meet a set of graph-theoretic constraints. The cognitive network approach to language permits the researcher to relax formal constraints, as long as the architecture is compatible with known properties of the mind and brain. This does not absolve the researcher from embedding the research programme in a set of constraints that limit possible theories. It means that the game is different: the constraints are not those of formal language theory; instead, the theory is obliged to be bound by findings from cognitive psychology and related areas. WG has a dependency theory of syntax because it is the syntactic theory that is compatible with the network theory of language and cognition. A dependency graph is a network with terminal nodes, so it is consistent with the idea that language is a cognitive network. For a cognitive theory, the issue comes down to two things: (i) is the structure parsable by the *human* parser? And (ii) is it learnable? To learn a grammar, the speaker/hearer will start small and build the grammar incrementally, learning words and dependencies, which can be learned from adjacent words. There is flexibility built in. Once learned a dependency can be subject to processing variations, even interruptions: *I was wondering—could you pass the salt please—whether you'd ever been to Italy.*

Therefore, for Hudson, mutual dependency is acceptable in syntax, because loops are found in ordinary relationships in cognition and loops are therefore learnable. In the rest of this paper, I argue that their learnability in the general case does not mean that loops should be tolerated in syntax. In particular, I argue that syntax is different from the rest of the cognitive network, for three reasons. The first is that while networks are non-directional, syntax is necessarily directional. The second is that syntax

<sup>2</sup>The lack of a 'root' and the single quotation marks around *Homer*, *Bart* and *Marge* show that this is a fragment of conceptual structure, not syntax.

necessarily involves terminal nodes, and therefore generating or parsing sentences always involves active searching for and retrieving information in an inheritance hierarchy. The third difference between syntactic dependencies and other cognitive relations is that dependency is a hierarchical relationship between head and dependent. These properties make dependencies conceptually richer than other network relations, and introduce the possibility of conflicting information in mutual dependency. In Section 2, I address these issues, then in Section 3 I re-examine Hudson's arguments for mutual dependency within the NP.

## 2 Why is syntax different?

The WG cognitive network is claimed to have the small-world and scale-free properties (Steyvers and Tenenbaum, 2005) found in semantic networks which are also found in dependency networks (Ferrer i Cancho and Solé, 2001; Ferrer i Cancho et al., 2004). This is unsurprising: dependencies are a particular subtype of grammatical relation, and grammatical relations are a subtype of the relations we find in cognition. However, as a subtype they are conceptually richer than the network relations that they inherit from. Dependencies are implicated in both form and meaning. They establish hierarchical relations among words. They are responsible for word order, combine forms, trigger the morphophonological facts of agreement, and they combine referents of words with the semantics of the heads of those words. We therefore need to be cautious about presuming that the gross similarities between dependencies and other cognitive relations mean that syntax is just like the wider cognitive network. Furthermore, there are different subtypes of dependency. Some, but not all, are associated with the landmark relation and are responsible for word order. And some, but not all, are valents.

Linearisation is self-evidently a key element of syntax so we can take it first. We can say *the dog* but not *\*dog the*; *in the kitchen* is part of English but *\*the in kitchen* is not. In this respect, syntax is different from the network structures elsewhere in cognition. But it gets more complicated. Take *La table que j'ai achetée est là*, 'The table I bought is there', where *achetée* agrees with *La table*. In French, perfect constructions with AVOIR do not usually trigger agreement between the participial complement and its subject, unlike those constructions with ÊTRE. There is no agreement in *Elles ont acheté la table*. However, when the participle's direct object is linearised before it, agreement is triggered between the direct object and the participle. This is not just a prescriptive rule: the agreement can also be heard as in *la lettre que j'ai écrite*. See also Cong and Liu (2014) and Liu et al. (2017) on linearisation.

Because syntax has to involve terminal nodes, it also involves both the retrieval of taxonomic information and a process of updating it. The rest of the cognitive network will also involve terminal nodes, because it will be constantly updated in the face of new information, from different types of perceptual information including speech perception, but syntax always involves terminal nodes because it involves the classification of utterances, which must be linearized, and which are related to the conceptual-intensional system of semantics. In WG, each word is an action with a time, place and speaker (=actor); each of these actions is linked to a concept in the conceptual-intensional/semantic part of the language network. The database of information that makes up the permanently stored mental network is a database of 'declarative knowledge, expressed as propositions' (Hudson, 1984, 2; emphasis original). This database of propositions licenses the utterances or allows the hearer to form inferences about the structure and determine a parse for what they hear, by exploiting default inheritance as they classify each utterance token. Word Grammar is a constraint-based theory and WG syntax is an interface between the stored database of propositional knowledge in the conceptual network, and the human behaviour of making and interpreting utterances. It involves a mental network sitting on top of a neural network, with undirected spreading activation in the network (Collins and Loftus, 1975). For mutual dependency to be possible, it has to be consistent with the constant searching for, and updating of, information that makes up real-time language use.

The hierarchical nature of syntax is baked into dependency analyses: dependency is a pairwise relationship between words where one is the head and the other is the dependent. Heads have certain properties: they are responsible for distribution, and also for the internal structure of a phrase.<sup>3</sup> For

---

<sup>3</sup>Within WG, Rosta (2006) has argued for factoring these notions out and contended that they are distinct in order to allow

example, in *De officiis* (Latin, ‘On duties’) *de* selects a noun, determines its position and governs its case (ablative). The preposition is clearly the head, *officiis* the dependent. In the case of mutual dependency, it is hard or impossible to know which word has chosen the other, which word is responsible for the form of the other, and which word is responsible for the (relative) position of the other. The selection and positional information is both hierarchical and linear: a word of type *x* requires a word of type *y* to its right or left and, in languages with the right kind of morphology, governs its form. The point about syntax being comprised of terminal nodes is that in WG this information has to be looked up in the inheritance hierarchy. It therefore has to be stored as a generalisation that words of type *x* have these properties. Furthermore, the hierarchical nature of syntax is involved at a depth of complexity: in *he<sub>i</sub> made the food for himself<sub>i</sub>*, the word *himself* is linked to *he* as a co-dependent (in some relevant sense) of the same head.

What do these facts mean for mutual dependency, given the analysis in Figure 3 and the observation that loops must exist in conceptual structure? It means we can accept that loops must exist in the symbolic network, while simultaneously requiring there to be further supporting theoretical evidence for mutual dependency in the syntax. We know that the nature of cognition places constraints on syntactic structure: there is a literature exploring how working memory constrains dependencies, particularly dependency distance, (Cong and Liu, 2014; Futrell, 2017; Futrell et al., 2020; Gibson, 1998; Gildea and Temperley, 2010; Liu, 2018; Liu et al., 2017) and it is appropriate to think about the relationship between working memory, language dynamics and language structure. In the case of dependency distance, Cong and Liu (2014, 605) cite spreading activation as key: ‘Given the small-world topology, whereby each pair of vertices can be generally connected by a short path, the loss of activation energy can be minimized and the success of retrieval is thus maximized.’ As they say, this gives rise to a preference for minimal dependency distance, which is a constraint on structure consequential upon the interaction of linear information flow and working memory limitations. What consequences do such facts have for the status of mutual dependency?

For there to be mutual dependency, it must be consistent with the design features of the cognitive network. It is with some of them: a mutual dependency relationship has a small-world structure which, following Cong and Liu (2014), should facilitate retrieval and minimise the loss of activation energy making the structure easily chunked and passed into longer-term storage. But this is not the whole story because there are other considerations. The specifically unique properties of syntax, its hierarchical, linear structure, make it possible for mutual dependency to be coherent to the extent that syntax fits a network topology, but not coherent in respect of hierarchy and linearity, or inheritance. The way to address this is by asking whether mutual dependency might cause problems for a WG analysis.

It does. A hierarchical, linear structure places constraints that do not apply in an ordinary network. In the case of the D+N construction, Hudson (2004, 32) writes, ‘Since we have seen that D and N depend on each other, either of them can be the head of the NP, and the choice can be left to the surrounding construction.’<sup>4</sup> Therefore, either word is the head, depending on the surrounding linguistic context. Within the phrase, with mutual dependency, both words are landmarks for the other, and select the other. These two consequences render the phrase effectively headless: D+N does not have a formal structure where we know that D, or N, is the head, responsible for the distribution of the phrase, selecting the other element, determining the linear position of the other, and governing the other element’s form.<sup>5</sup>

There are two possible consequences for such a theory. The first is that the headedness of D+N is unknown until the construction containing the D+N string resolves it, in which case D+N is unique among the major, frequently attested, constructions of English in being both headless and constructional. This would mean that the speaker/hearer would not know which word was the head until the constructional context determined it, creating a problem for eager parsing in the word-by-word approach to dependency parsing of Covington (2001). Here, as each word is encountered, it is attached to a classification which

---

for a range of phenomena such as pied-piping. For Rosta, in the grammar of English, the structure of a phrase is mainly decided by evidence such as word order and ellipsis. Distribution concerns the positions where a phrase may occur.

<sup>4</sup>The thin entering wedge for this position is Hudson’s treatment of determiners as (transitive) pronouns, a position he adopts rather than Postal’s analysis that pronouns are determiners. Given that pronouns are in turn nouns, both elements in D+N are nouns, and therefore hypothetically either may serve as head.

<sup>5</sup>Mutual selection is not inherently a problem; it is determining the head for distribution and structure that is. Mutual selection is indirectly a challenge, however, because it contributes to the problem of identifying the head.

tells the speaker/hearer what to expect about its behaviour. Perhaps at best the internal structure of a mutual dependency phrase could be resolved. But on this approach, neither word has an incoming dependency relating it to its head until that is determined by the containing construction—which could be to the right of the D+N string at issue. This means that the D+N string has to be held in memory until its head can be determined. But processing is rapid, and anticipatory, and has to happen in the context of Christiansen and Chater (2016)'s 'Now-or-Never bottleneck', which is a working-memory constraint on linguistic production. Due to the fleeting nature of memory, the brain has to compress and recode linguistic information fast and efficiently, across all of the levels of grammar, otherwise the information is lost. Anticipatory parsing is a consequence of the bottleneck; it requires linguistic units to be chunked and then passed up to longer-term storage as the information flow moves along. It is not possible if we do not know which word is the head until the wider context tells us.

The other possible consequence is that there exist two propositions, 'D is head' and 'N is head', and the speaker/hearer has to resolve them. The problem with the two proposition approach is that it involves a contradiction. Like everything else in the cognitive ecology, headedness is learnt on the basis of experience. To learn mutual headedness in syntax would need a model where it was not contradictory. A contradiction can only be resolved by stipulation, otherwise it gives rise to a failure of structure such as with the Nixon diamond in multiple inheritance (Touretzky, 1986) which, Hudson (2000) argues, gives rise to ungrammaticality, and accounts for the lexical gap of *amn't*. Grammars are (by and large) regular, coherent and learnable, which requires simple structures that facilitate rapid structure building in the online work of parsing and production. If contradictions render lexical gaps, this contradiction should render a mutual dependency analysis of D+N impossible, and in turn, if a mutual dependency analysis were required, there should be a constructional gap: \*D+N.

There are two possible routes out of this pair of problems. One is to adopt the approach of Rosta (2006) mentioned in Footnote 3 and to factor out different dimensions of headedness. As long as only one word is the structural head, and only one the distributional head, this approach will work for D+N. Another is to allow mutual syntactic relationships, but to make only one of those relations a dependency, with the other relation carrying a depleted degree of syntactic information. Such an approach is permitted by the network topology and consistent with Hudson's theory of landmarks for word order. Moreover, by making only one of the words concerned responsible for distribution and non-conflicting aspects of the internal structure of a phrase, this approach avoids the disadvantages of mutual dependency. However, the evidence is clearly that mutual dependency is a problem for Word Grammar, for all that other kinds of loop in the syntax need not be. I now turn to the Hudson/van Langendonck arguments about headedness of the D+N construction. I argue that the arguments Hudson offers in favour of N do not have sufficient force that we are obliged to adopt a mutual dependency analysis of this construction.

### 3 The head of D+N

In this section, I concentrate on the arguments that Hudson (2004) adduces in favour of mutual dependency of D+N. I do not address his earlier arguments that D is the head, nor do I address other arguments in the literature about the headedness of D+N such as those in Osborne (2021). My concern in this section is in establishing whether the five arguments Hudson (2004) presents in favour of N as head are sufficient reasons to adopt a mutual-dependency analysis of D+N. The shape of the argument is that if his evidence that N is (also) the head can be adequately challenged or is inconclusive, there is no reason from the data to adopt mutual dependency.

We can begin by thinking about what determiners do. In (3), there are a number of nouns of different kinds, serving as the subject of the main verb, with various constraints shown.

- (3) a. Ovid was banished to Tomis.
- b. She is in the sitting room
- c. Water flowed out of the kitchen door.
- d. Students poured out of the classroom.
- e. \*Dog trotted down the street.
- f. \*The my dog trotted down the street.

- g. John's dog trotted down the street.

In (3) we see that proper nouns, pronouns, mass nouns and plural count nouns can all occur without a determiner (3a-d), but singular count nouns cannot (3e). The example in (3f) tells us that it is not possible to have chains of determiners in English. And (3g) tells us that possessives also make a singular count noun grammatical. These restrictions are language specific: for example, proper names in Ancient Greek have articles and Latin allowed examples such as (3e). The theory, then, must be a parochial account of English. Finally, the subjects in (3a-d) are all referential. Proper nouns and pronouns are necessarily referential, and the evidence of (3c-d) is that mass nouns and bare plurals are weakly existential, but bare singular count nouns are not. One of the things that determiners do is to fit singular count nouns up to refer. Another is that they make the reference of other nouns more precise (or determine their reference).

Hudson (2004) argues that mutual dependency is necessary in the analysis of Determiner+Noun strings because of various arguments in favour of N as head published in Van Langendonck (1994). Given that Hudson (2004) continues to find his own earlier arguments that the Determiner is the head compelling, he concludes that there is a state of mutual dependency between the determiner and the noun, and that the headedness of the construction is moot. The most straightforward way to tackle whether mutual dependency is necessary in a WG analysis of D+N is to assume that Hudson's earlier analysis will work, and to explore how compelling the additional arguments for N as head turn out to be on a second look. If it is possible to dispose of the new arguments in Hudson (2004) then it is possible for the theory to dispense with mutual dependency in this area of grammar by reverting to the earlier analysis. Hudson (2004) relies on Van Langendonck (1994) for arguments that N is head as well as Osborne (2003) and Huddleston and Pullum (2002).

The main arguments in Hudson (2004) that D depends on N are: (i) NPs as adjuncts; (ii) Possessives (3g); (iii) the need for determiners (3d); (iv) the single determiner constraint (3e); (v) facts about extraposition. I take these in turn.

### NPs as adjuncts

The relevant examples are shown in (4), taken with the quotations from Hudson (2004, 11); he takes this as the most important set of facts in Van Langendonck (1994).

- (4) a. I saw him *this morning*  
b. It's best to do it *my way*  
c. Put it *this side of the line*

Hudson's claim is that the 'NPs that can be used in this way are defined exclusively in terms of their N; the D is more or less irrelevant, being freely selected according to the normal rules.' There are two further key restrictions: the italicized NPs in (4) cannot be replaced by a personal pronoun or *this* without a complement, and '[A]lthough all the eligible nouns all refer to times, places and manners, they are also lexically quite restricted': for example, it is possible to use WAY in these constructions but not its (near) synonym MANNER. See (5).

- (5) a. \*I saw him it.  
b. \*It's best to do it mine  
c. \*Put it this.  
d. I did it the usual way/\*manner.

A further restriction is clear with NP time adjuncts: we can say *I saw him this morning* but not *\*I saw him this party* even though PARTY can refer to a time in an expression such as *I saw him before the party*. However, it is unclear whether this is a separate lexical restriction, or whether there is a semantic generalization. Perhaps PARTY is excluded because its basic semantics is to refer to an event, and it refers to the time of the event by metaphorical extension. So how should we account for these facts?

The examples are time, place and manner adjuncts and the key facts about adjuncts are that they are not selected; they define their own semantic relation to their head; and they 'reverse unify' with their heads. They are syntactic dependents, but in the semantics they take their heads as their arguments.

Hudson claims that these D+N constructions should be treated as having the N as the head because it appears to decide their ability to occur as adjuncts. However, there is a further set of examples that he introduces that make the situation more complicated. Adjunct D+N patterns also place restrictions on the determiners that occur in them, (6).

- (6) a. He did it this morning/\*the morning/\*a morning.  
b. He did it this way/\*the way/\*a way.

As (6) shows, THE and A cannot occur in these adjunct NP constructions, unless the N is further modified: *he did it the right way* and *he did it the same morning* are both fine. The restrictions are also related to the noun in the construction: as Hudson (2004, 12) points out there is variation among the nouns that can occur in adjunct NP constructions in terms of which determiners they occur with.

- (7) a. I'll do it in my (own) time  
b. I'll do it on my day.  
c. \*I'll do it my time/day.

Although TIME and DAY can occur in NP adjuncts (*I saw him this time/that day*), and can occur with possessives within PP adjuncts, they cannot occur with a possessive when they function as NP adjuncts.

The final set of facts discussed by Hudson has to do with relative clauses. He notes that the restrictions on nouns in adjunct NPs survive relativisation, where the noun is the external head of the relative clause, as in (8) (Hudson, 2004, 13). In the examples, 'way is possible but *manner* is not, and *time* is possible but *point(in time)* is not.'

- (8) a. The way/\*manner he did it shocked us.  
b. I remember the time/\*point he did it.

Hudson (2004, 13) uses the relative clause structure in (8) as an argument that the D is irrelevant. However, there is a complication that Hudson does not discuss: although (9a) is fine, none of the examples in (9) are acceptable.

- (9) \*My/\*this/\*that way he did it shocked us.

To summarise the restrictions: only nouns with the right semantics can appear in these adjunct constructions; and only certain nouns can appear in the construction with their apparent synonyms being excluded; there are environmentally conditioned restrictions on the determiners that occur in these constructions.

It is worth noting that the restriction in (6) co-varies with whether the noun is modified or not. The relative clause restriction in (8) provides some evidence, but also they restrict modification in examples such as *\*I saw him this inconvenient time*. However, *I saw him that terrible day* is fine.

It is not obvious that these facts argue against the traditional WG analysis of D+N, with D as head. The restrictions on nouns such as MANNER and different determiners resemble those with *kick the bucket*.

- (10) a. He kicked the bucket/\*a bucket/\*this bucket/\*some bucket.  
b. \*He kicked the pail/scuttle/pitcher.

Just as in idioms, there are restrictions on both N and D in the NP adjuncts: it appears that NP adjuncts are idioms. How should we treat idioms? There is a WG analysis in Gisborne (2020, 44-55) which we could adopt. The analysis relies on WG's default inheritance architecture: in the case of *kick the bucket*, there is a special subtype of the lexeme KICK, which selects a special subtype of THE which in turn selects a special subtype of BUCKET. That is to say that the whole string is defined as a regular collocation, with a particular meaning, but the normal syntax of the phrase is maintained with *the* still the head. A similar analysis of NP time and manner adjuncts would help capture the degrees of irregularity we see in these examples. We know, both from the discussion in Gisborne (2020, 44-55) and also from Nunberg et al. (1994) that there is potentially a great deal of variation within the expression of different idioms.



Given the apparently arbitrary restrictions on NP adjuncts, they invite a similar analysis where the various lexical restrictions—on both D and N—are treated as reflexes of the construction’s idiomaticity. The appropriate analysis would then be to treat the different variants as subtypes (sublexemes in WG’s terms) of an NP time or manner adjunct type in the inheritance hierarchy.

The place adjuncts in particular argue in favour of this analysis. *Put it this side of the line* requires the mention of a (semantic) landmark relative to the line: *Put it my/your/his/this/that side of the line* are all fine, but *\*Put it the side of the line* is not because the definition of where, relative to the line, is relevant to the definition of place that the adjunct is contributing. This also shows that there are degrees of idiomaticity in this area of grammar, which is consistent with what Gisborne notes for idioms more generally, and is also by and large compatible with Nunberg et al. (1994)

The one argument we still need to discuss is the argument from the interaction of adjunct NPs and bare relative clauses in examples such as (8). Hudson argues that as N is the antecedent of a relative clause, not NP, the structure of (8)(a) is *The [way he did it] shocked us* rather than *[[The way] he did it] shocked us*. The analysis that the relative clause depends on N (or its analogue in phrasal theories) is standard since Partee (1975); the claim that N is the antecedent of the gap is also widely adopted, see e.g. Huddleston and Pullum (2002, 1037). The reasoning is that the determiner determines the whole of the syntagm [N+relative clause], and so D+N cannot be the antecedent of the gap. However, although the attachment of the relative clause is clear, it is not clear that it tells us about the filler-gap relationship, because there is a conflict of facts: bare singular count nouns cannot occur as the dependents where gaps are found. This is easier to see with argument gaps than adjunct gaps, because their positions are fixed. For example, *The party he got drunk at \_\_ was yesterday* requires a determined count noun to fill the gap because *\*He got drunk at party* is ungrammatical. This implies that the antecedent of the gap is *the party* even though the relative clause *he got drunk at \_\_* modifies *party*, not the whole of the NP: the preposition AT cannot occur with a bare count noun. On the other hand, Sauerland (1998, 65ff.) presents reconstruction evidence for a relationship between the head noun and the gap. The best conclusion is that bare relatives are a topic for further research, not a knock-down argument for N as head.

### Possessives

Hudson presents two different arguments from Van Langendonck concerning possessives. The first is evidence from Dutch, which is only indirectly relevant, because Dutch is not English and the grammar of English NPs involves parochial facts about English. See (11).

- (11) a. Moeders jurk ‘mother’s dress’  
b. Peters moeders buren ‘Peter’s mother’s neighbours’.

The argument is that (11a) is a hyponym of *jurk* ‘dress’, so this is the head; however the same is true for a phrase such as *the dog* which is a hyponym of *dog*, where previously Hudson has argued that *the* is the head. Here, I think, we can take a different approach. The key fact is that in Dutch (and German for that matter) a singular count noun cannot occur on its own as in English, but it is not only rendered grammatical by a determiner: a genitive noun, which cannot co-occur with a determiner, can also make it grammatical. Van Langendonck and Hudson’s argument is that the genitive noun is dependent, and therefore the determiner must be dependent.

Perhaps this is so. But first, English does not have case, except vestigially on personal pronouns, so the arguments are not directly relevant and the comparison has to be appropriately set up. In fact, there is a comparison which obviates the need to take N as head, which we can find by thinking about what determiners do. In an article developing a WG theory of the diachrony of the English definite article, Gisborne (2012) argues that the article is a quantifier expressing both existential and universal quantification over the noun, following arguments in Russell (1905) and Neale (1990). If the purpose of articles is to quantify over nouns, then by extension this is the general purpose of determiners, because the treatment of definite articles extends, as Gisborne discusses, to other definite elements, and because there are already determiners whose role is to quantify. The indefinite article simply provides existential quantification. Philippi (1997) provides interesting evidence for this position in a discussion of the emergence of the article system of German, and also provides further evidence that case can have existential force, thereby

quantifying existentially over the noun—a suggestion which is supported by differential case marking. If this is right, we can account for the Dutch data in (11) with a simple observation: the genitive noun is relational. The genitive case gives existential quantification over both the noun that it is attached to and the noun that the genitive relates to. There is no scope for double determination (or quantification) and so the similarity between the Dutch genitive in (11) and English possessive 's is captured.

There are further arguments about the synonymy of *the old man's hat* and *the hat of the old man*. Hudson (2004, 16) addresses those arguments. Van Langendonck's main argument is from the idiom *to pull someone's leg*, meaning to tease someone. The observation is that in the idiom, the possessor is required in both variants, and *\*She pulled the leg* is ungrammatical in the idiomatic sense (although it is perfectly grammatical if it is taken literally, and said about the chicken on the table). This argument has little force, because it concerns the structure of an idiom. In much the same way as we can omit neither the head nor the direct object of *She kicked the bucket* nor change either of them, we cannot rework the structure of this idiom without it losing its idiomatic meaning.

### **The need for determiners**

Hudson presents a number of arguments from the need for determiners. The argument is that if a word is required by another word then it is the dependent of that word, as in the case of valency. Singular count nouns require a determiner to be able to occur in argument positions; without one, if they occur in an argument position they are forced to a mass interpretation: *Dog was all over the road*. Although this looks as if D is required by N, we can take the argument in the previous section and rethink the facts. Let us assume that determiners quantify over nouns. How are nouns quantified over without a determiner? The examples in (3) show that there are different ways in which a noun might be quantified over. Proper nouns and pronouns are inherently referential. Possessives are referential because 's is inherently definite. Plurals are weakly existential, because plural marking asserts the existence of more than one of the entities denoted by the noun. Mass nouns are the most difficult to describe in these terms. In examples such as *Water is necessary for life* they are weakly generic. The example in (3)(c) is weakly existential: *Water flowed out of the kitchen door* asserts the existence of the water and it is upward entailing. Generics are not upward entailing, on the other hand. However, it is also different from *Some water flowed out of the kitchen door*, which implies a finite mass of water. I think that this property of mass nouns follows from the nature of massness. Mass nouns refer cumulatively: if I show you a heap of rice, I can say, 'This is rice'; I can then show you another heap and say the same, which I can also say of both heaps. This shows that mass nouns by default presume the existence of the stuff that they denote, and they do not require external quantification, unless we quantify over them partitively.

From this we can conclude that singular count nouns are the only subtype of noun requiring some linguistic formative to assert their existence. In case-marked languages, case itself can do this, as is indicated by differential case marking. For example, in Finnish partitive case gives rise to an indefinite interpretation whereas accusative objects are interpreted definitely. In a language such as English, on the other hand, in the absence of case a singular count noun cannot occur unless it is quantified by a determiner. As the determiner quantifies over the singular count noun, it must be the head.

### **The single determiner constraint**

Only one determiner is permitted in D+N structures. Because heads can only select one of a given type of dependent, Hudson sees this as an argument that D is a complement of N: the argument is that there is a single 'slot' for a determiner in the grammar of a noun, in much the same way as it is only possible for a verb to have a single subject or a single direct object, and so once it is filled no other element can occur in that position. This suggestion has the advantage of capturing the generalisation about Dutch genitives discussed above. But if we exclude examples of nominal modification, such as *the boy actor*, only one noun is permitted which is why *\*the dog cat* is ungrammatical, and to have two nouns determined by a single determiner, they have to be coordinated: *the dog and cat*.

But the constraint does not only apply to dependents. It also applies to heads: it is also only possible to have one head in a given structure. We can analogise from auxiliary verbs. In a string of English auxiliary verbs, it is only possible to have one finite verb: *She may have gone* is grammatical; *\*She may has goes* is not. This is simply a matter of selection. Each auxiliary selects the form of its complement.

Likewise, determiners. It is part of the grammar of THE that its complement must be a common noun and not another determiner. Other quantifying expressions such as ALL will permit a definite determiner, on the other hand: *all my teachers*; *all those dogs*; *all the cakes*. This constraint can be argued either way, and it is not a knock-down argument for N as head.

### Extrapolation

The final constraint which Hudson (2004, 20) takes as evidence for N as head is extrapolation. The constraint is shown in the example in (12). The argument is that it is only possible to extrapolate dependents of nouns which themselves depend directly on the main verb.

- (12) a. People [who have been waiting ten years] are still on the list.  
b. People are still on the list [who have been waiting ten years]  
c. \*Names of people are still on the list [who have been waiting ten years].

However, the generalisation in (12) is not the full description. In (13) the extrapolation is possible, even though the noun does not depend directly on the verb.

- (13) a. All of the people [that have been waiting ten years] are still on the list.  
b. All of the people are still on the list [that have been waiting ten years].

The conclusion is that it is possible to extrapolate dependents of nouns which do not depend directly on the main verb, and that the data in (12) need a separate explanation. The evidence in (13) does not only undermine the argument that (12) is claimed to demonstrate: it also shows that it is possible to extract a dependent of N when N is itself quantified over. If the earlier claims that D quantifies over N are correct, then (13) suggests that extrapolation is fine out of a quantified N, and also that D is a quantifier.

## 4 Conclusions

Primarily because of the problem it causes in identifying the head, and the problems that this brings about for processing, I have argued that WG should reject mutual dependency, despite its network architecture and cognitive basis. This is not an argument that there should be no syntactic loops; it is an argument that if loops are to feature in the syntax, they should not be found among the dependencies which are relevant to distributional structure. There should be no mutual dependency with each word (potentially) the head of the other and the landmark of the other. I have shown that the arguments of Van Langendonck and others which Hudson (2004) draws on in the development of his mutual dependency account of D+N can all be addressed, thereby obviating the need to assume mutual dependency. Where these arguments are inconclusive, there is no need to adopt mutual dependency. Where I have offered alternative analyses, these are arguments against N as head, in which case mutual dependency should not be adopted. In some cases, I have relied on arguments that D is a quantifier and that it quantifies over N: such an analysis is inherently compatible with the D as head analysis. The advantage of the present account is that it leaves the syntax of D+N asymmetrical and single headed and therefore consistent with the most essential premise of a dependency grammar. It is also more consistent with a word-by-word theory of dependency parsing, and does not involve a learnability problem by introducing a contradiction. I have, however, left open for future research the other mutual dependency structures that Hudson (1990) introduces.

## Acknowledgements

I am grateful to Dick Hudson for comments on the first draft; Doug Arnold, Ronnie Cann, Adam Przepiórkowski, Geoff Pullum and Rob Truswell for discussing points of detail (they bear no responsibility at all for what I have done with their help); and three referees for very useful comments which have improved the paper (with apologies that space limitations prevented me from incorporating all of their suggestions). I am, of course, responsible for remaining errors and infelicities.

## References

- John M Anderson. 2006. *Modern Grammars of Case*. Oxford University Press, Oxford.
- Miguel Ballesteros and Joakim Nivre. 2013. Going to the Roots of Dependency Parsing. *Computational Linguistics*, 39(1):5–13, March.
- Morten H. Christiansen and Nick Chater. 2016. The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39:e62.
- Allan M Collins and Elizabeth Loftus. 1975. A spreading-activation theory of semantic processing. *Psychological Review*, 82(6):407–428.
- Jin Cong and Haitao Liu. 2014. Approaching human language with complex networks. *Physics of Life Reviews*, 11(4):598–618, December.
- Michael Covington. 2001. A Fundamental Algorithm for Dependency Parsing. In John A. Miller and Jeffrey W. Smith, editors, *Proceedings of the 39th Annual ACM Southeast Conference*, pages 95–102. Association for Computing Machinery.
- Chet Creider and Richard Hudson. 2006. Case agreement in Ancient Greek: Implications for a theory of covert elements. In Kensei Sugayama and Richard Hudson, editors, *Word Grammar: New Perspectives on a Theory of Language Structure*, pages 33–53. Continuum, London.
- Ramon Ferrer i Cancho and Richard V. Solé. 2001. The small world of human language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1482):2261–2265, November.
- Ramon Ferrer i Cancho, Ricard V. Solé, and Reinhard Köhler. 2004. Patterns in syntactic dependency networks. *Physical Review E*, 69(5):051915, May.
- Richard Futrell, Roger P. Levy, and Edward Gibson. 2020. Dependency locality as an explanatory principle for word order. *Language*, 96(2):371–412.
- Richard Landy Jones Futrell. 2017. *Memory and Locality in Natural Language*. Ph.D. thesis, MIT, Cambridge MA.
- Haim Gaifman. 1965. Dependency systems and phrase-structure systems. *Information and Control*, 8:304–337.
- Edward Gibson. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68:1–76.
- Daniel Gildea and David Temperley. 2010. Do Grammars Minimize Dependency Length? *Cognitive Science*, 34(2):286–310, March.
- Nikolas Gisborne. 2012. The semantics of definite expressions and the grammaticalization of THE. *Studies in Language*, 36(3):603–644.
- Nikolas Gisborne. 2020. *Ten Lectures on Events in a Network Theory of Language*. Brill, Leiden.
- David G. Hays. 1964. Dependency theory: A formalism and some observations. *Language*, 40(4):511–525.
- Rodney Huddleston and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge.
- Richard Hudson. 1984. *Word Grammar*. Blackwell Oxford.
- Richard Hudson. 1990. *English Word Grammar*. Blackwell, Oxford.
- Richard Hudson. 2000. I amn’t. *Language*, pages 297–323.
- Richard Hudson. 2004. Are determiners heads? *Functions of Language*, 11(1):7–42.
- Richard Hudson. 2007. *Language Networks: Towards a New Word Grammar*. Oxford University Press, Oxford.
- Richard Hudson. 2010. *An Introduction to Word Grammar*. Cambridge University Press, Cambridge.
- Timo Järvinen and Pasi Tapanainen. 1998. Towards an implementable dependency grammar. In *Processing of Dependency-Based Grammars*.

- Haitao Liu, Chunshan Xu, and Junying Liang. 2017. Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, 21:171–193, July.
- Haitao Liu. 2018. Language as a human-driven complex adaptive system. *Physics of Life Reviews*, 26–27:149–151, November.
- Ryan McDonald and Joakim Nivre. 2011. Analyzing and Integrating Dependency Parsers. *Computational Linguistics*, 37(1):197–230, March.
- Igor A. Mel'čuk. 1988. *Dependency Syntax: Theory and Practice*. SUNY press, Albany, New York.
- Igor A. Mel'čuk. 2014. Dependency in language. In Kim Gerdes, Eva Hajičová, and Leo Wanner, editors, *Dependency Linguistics: Recent Advances in Linguistic Theory Using Dependency Structures*, pages 1–32. Benjamins, Amsterdam/Philadelphia.
- Stephen Neale. 1990. *Descriptions*. MIT Press, Cambridge MA.
- Geoffrey Nunberg, Ivan Sag, and Thomas Wasow. 1994. Idioms. *Language*, 70:491–538.
- Timothy Osborne. 2003. *The Third Dimension: A Dependency Grammar Theory of Coordination for English and German*. Ph.D. thesis, Pennsylvania State University, State College PA.
- Timothy Osborne. 2019. *A Dependency Grammar of English: An Introduction and Beyond*. Benjamins, Amsterdam.
- Timothy Osborne. 2021. NPs, not DPs: The NP vs. DP debate in the context of dependency grammar. *Acta Linguistica Academica*, 68(3):274–317.
- Barbara Partee. 1975. Montague Grammar and Transformational Grammar. *Linguistic Inquiry*, 6(2):203–300.
- Julia Philippi. 1997. The rise of the article in Germanic languages. In Ans Van Kemenade and Nigel Vincent, editors, *Parameters of Morphosyntactic Change*, pages 62–93. Cambridge University Press, Cambridge.
- Jane J. Robinson. 1970. Dependency structure and transformational rules. *Language*, 46(2):259–285.
- Andrew Rosta. 2006. Structural and distributional heads. In Kensei Sugayama and Richard Hudson, editors, *Word Grammar: New Perspectives on a Theory of Language Structure*, pages 171–203. Continuum, London.
- Bertrand Russell. 1905. On denoting. *Mind*, 14(56):479–493.
- Uli Sauerland. 1998. *The Meaning of Chains*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Reidel, Dordrecht.
- Mark Steyvers and Joshua B. Tenenbaum. 2005. The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, 29(1):41–78, January.
- Lucien Tesnière. 1953. *Esquisse d'une Syntaxe Structurale*. Librairie C. Klincksieck, Paris.
- Lucien Tesnière. 1959. *Éléments de Syntaxe Structurale*. Librairie C. Klincksieck, Paris.
- David S Touretzky. 1986. *The Mathematics of Inheritance Systems*, volume 8. Morgan Kaufmann, Los Altos, CA.
- Willy Van Langendonck. 1994. Determiners as Heads? *Cognitive Linguistics*, 5(3):243–259.