English Language and Linguistics 16.3: 385–401. © Cambridge University Press 2012
doi:10.1017/S1360674312000160

Branching direction in recursive structures[1]

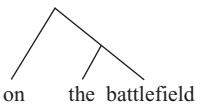
THOMAS BERG
University of Hamburg
(Received 28 April 2011; revised 10 January 2012)

English makes regular use of a number of recursive structures spanning the syntax–lexicon continuum. While NPs with recursive relative clauses occupy the syntactic end, nominal compounds are located at the lexical end. In between these extremes we find NPs with recursive periphrastic genitives (towards the syntactic end) and NPs with recursive Saxon genitives (towards the lexical end). This study presents a comparative analysis of the branching direction preferences in these recursive structures. The empirical focus is on double *of*-genitives, which exhibit an overwhelming predilection for right-branching. This contrasts sharply with the double Saxon genitives, which gravitate towards left-branching. The branching direction decision is argued to be under the sway of several distinct factors: a syntactic factor controlling the alternative between leftward and rightward expansion; a lexical factor regulating the idiomatization of a given pair of elements; and a processing factor geared towards preventing garden path effects. Furthermore, branching direction is determined by listeners' desire to minimize constituent recognition domains. Taken together, these factors are held accountable for the varying branching direction biases found in the different types of NP.

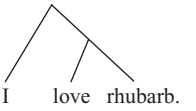
1 Introduction

The constituent structure of words and phrases is standardly determined on the basis of syntactic and semantic criteria. At least three different types may be distinguished depending on the relative importance of these two sets of criteria. In the first type, one syntactic unit always gives rise to the same constituent structure irrespective of its meaning, i.e. the lexical items with which its slots are filled. Classical examples include the right-branching analysis of English PPs and sentences with SVO order, as shown in (1) and (2), respectively. There is general consensus that semantics is irrelevant in these cases.

(1)



(2)



[1] It is always gratifying to acknowledge the unflagging service I obtained from Marion Neubauer. As in several other projects, she handled the data collection task so diligently. I am also thankful to Melissa Neitzel, Nigel Isle and Craig Davis for their assistance with the analysis of the data as well as to an anonymous reviewer for helpful hints on how to clarify the presentation.

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

The second type covers semantic ambiguity. One and the same unit gives rise to different structural analyses which reflect different readings. Thus, semantics overrides syntax. Two examples, one lexical and the other syntactic, may serve as illustrations. The word *undoable* possesses a right-branching structure when it means 'unable to be done', but a left-branching structure when it means 'able to be undone' (Haspelmath 2002: 93). The NP *more exciting ideas* is right-branching when *more* modifies *ideas* (as in (3)), but left-branching when *more* modifies *exciting* (as in (4)) (Burton-Roberts 1986).

(3)

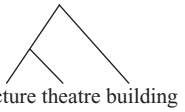


(4)

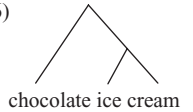


For expository reasons, the third type will be divided into two subtypes according to the linguistic level being targeted. Type 3A involves one and the same structural unit at the lexical level whose slots may be filled by different morphological units. It is the meaning of the entire word which generates a particular constituent structure. Because different words with the same morphological structure carry different meanings, they require different structural analyses. Consider prefix-stem-suffix patterns. The contrast between left- and right-branching is exemplified by Bolinger & Sears's (1981: 79) word pair *disgraceful* vs *ungraceful*. The semantic criterion imposes a left-branching structure on *disgraceful* ('full of disgrace'), but a right-branching structure on *ungraceful* ('not graceful') (for a more thorough analysis of *disgraceful*, see Plank 1985; for a lexicon-wide analysis of prefix-stem-suffix structures, see Berg 2003). Another case in point is three-noun compounds which admit of alternative branching directions. Whereas *lecture theatre building* requires a left-branching structure (see (5)), the right-branching analysis is appropriate for *chocolate ice cream* (see (6)).[2]

(5)



(6)



Type 3B is type 3A transferred to the syntactic level. It is not unreasonable to assume that the structural freedom observed at the lexical level in type 3A also shows at the syntactic level. Enriching syntactic patterns with different lexical items generates disparate meanings which may then form the basis of differing branching direction decisions.

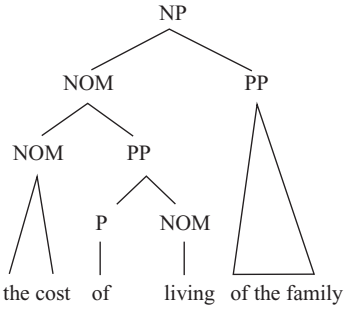
[2] Note that one and the same compound may also yield different constituent structures (e.g. *student drama society*) and therefore belong to the second type above.

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

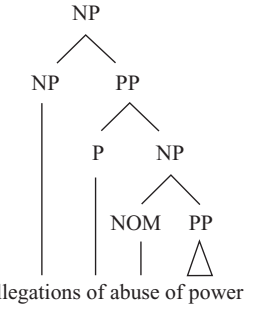
It is a noteworthy fact that relatively little is known about the constituent structure of type 3B cases. One wonders whether these exhibit a predominant branching direction bias and which factors tip the scales in favour of one or the other branching direction. Moreover, do different syntactic units entail the same or different structural configurations? It would seem natural to expect similar structural patterns to be subject to the same branching direction bias.

These questions will be addressed by focusing on recursive patterns, in particular NPs with periphrastic *of*-genitives in English. Like the so-called Saxon genitive, the *of*-genitive may be used recursively. With one level of recursion, this process leads to a complex NP consisting of a head noun postmodified by two PPs with P = *of*. As an assumed member of type 3B, this structure may be either left-branching or right-branching, as schematically represented in (7)–(8). The detailed discussion of concrete examples will be deferred to section 4.

(7) left-branching



(8) right-branching



The critical issue is the way in which the lower PP is integrated into the overall structure. If it is indirectly dominated by the higher PP, it creates a right-branching structure, as in (8); if, however, it modifies a complex nominal with an embedded PP, it is part of a left-branching structure, as in (7). Thus, left-branching arises from the expansion of the (left-hand) NOM, right-branching from the expansion of the (right-hand) PP at the level of immediate constituents.

Note that there is an asymmetry between (7) and (8). Whereas *of*-genitives allow consistent right-branching, they do not permit consistent left-branching. This is because PPs invariably occur to the right of their head nouns in English.

The major empirical issue with which the present article is concerned is which of the two branching directions is the predominant one and what the strength of this bias is. On a more theoretical level, it will be of interest to identify the factors that underlie this frequency distribution. In a wider perspective, the structural characteristics of the periphrastic genitive will be compared to those of other recursive patterns, with an eye to developing a general account of branching direction preferences. However, we will begin with a brief look at what is known about a related structural pattern, which will serve as the basis for deriving predictions pertinent to *of*-genitives.

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

2 The internal structure of recursive Saxon genitives

The closest semantic neighbour of the *of*-genitive is the Saxon genitive. The two main formal differences between them involve the marking of the possessive relationship by morphological or lexical means and the order of possessor and possession. In a previous study by Berg (2011), recursive Saxon genitives were found to be susceptible to both left- and right-branching, as exemplified in (9) and (10) from the Corpus of Contemporary American English.

(9) the director of the British Columbia *Women's Hospital's milk bank* in Vancouver
(10) the museum is more interested in being a primary part of *the city's visitor's bureau* by showing off a 'blockbuster' painting . . .

The fact that the Women's Hospital in Vancouver possesses a milk bank makes (9) a left-branching structure, while the fact that the city runs a visitor's bureau makes (10) a right-branching structure.

Berg (2011) observes a significant left-branching bias in recursive Saxon genitives. There are 70 per cent left-branching cases and 30 per cent right-branching cases. This result contrasts sharply with the general right-branching nature of the English language (Berg 2012). Berg (2011) proposes a perceptually driven frequency account for this asymmetry. The starting point is the left-to-right nature of speech. As listeners take in the acoustic signal, they immediately begin processing rather than wait until all of the information is available. When they hear, let us say, the beginning of a Saxon genitive (e.g. *Daniel's*), they build up a structural frame of the type 'Possessor*'s* Possession', thereby creating a link between the first noun (*Daniel*) and the upcoming NP. At this point, they do not normally know whether the Saxon genitive is recursive or non-recursive. However, they heuristically assume a possessor noun to be followed by a single possession noun for the simple reason that this non-recursive construction occurs far more frequently than the recursive construction in their language.

This marked frequency contrast combined with the principle of immediate processing has a noteworthy consequence for the ease with which listeners may expand a non-recursive into a recursive Saxon genitive in on-line processing. In a left-branching structure such as (7), the original hypothesis that the construction consists of a possessor and a possession continues to be valid. The close connection between the first two nouns remains appropriate even when listeners encounter a second genitive '*s* and a third noun. Thus, it is a straightforward task to integrate a non-recursive two-noun structure into a recursive left-branching three-noun structure in on-line processing.

By contrast, it is cognitively much more costly to turn a non-recursive construction into a recursive right-branching construction such as (8). The original analysis according to which the first two nouns are sister constituents has to be abandoned. Listeners have to start all over again and erect a new structure in which the major cut occurs between the first and the second noun and in which the second and the third noun are immediate constituents. This situation is reminiscent of the garden path effect observed in syntactic processing (e.g. Frazier & Rayner 1982; Rayner,

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

Carlson & Frazier 1983; Hawkins 2004), whereby listeners are misled into constructing inappropriate syntactic representations. For example, in the sentence *You know the answer is wrong*, the NP *the answer* is initially assigned object status although it is the subject of the embedded clause.

The claim is, then, that language structure is sensitive to processing difficulty. A structure which is difficult to process, or which is difficult to expand into another structure, tends to be used at a reduced frequency so as to avoid elevated mental cost or inefficient processing. This sensitivity discourages the occurrence of right-branching, and encourages the occurrence of left-branching NPs with double Saxon genitives, hence the unexpected left-branching bias in the English data. This bias is introduced by recursion insofar as listeners do not know beforehand that a recursive structure is coming up and therefore initially interpret the input as a non-recursive structure by treating the first two constituents as a unit.

The Garden Path Theory makes a clear prediction regarding the internal structure of double *of*-genitives. Since these are also created through recursion, listeners would be expected to bind the nominal and the first PP together and thereby make the expansion into a left-branching recursive structure a simpler task than the expansion into a right-branching one. The logic of the argument is entirely parallel to that of the double Saxon genitive. Because NPs with simple *of*-genitives are much more likely to occur than NPs with recursive *of*-genitives, listeners are prone to link up the nominal and the first PP. Note that the prediction of a left-branching preference holds irrespective of the order of possessor and possession. The first two constituents are predicted to be associated with each other in either case.

It is remarkable that the opposite expectation was aired by Plank (2002). He suspects that right-branching NPs with recursive *of*-genitives are more common than left-branching NPs. However, he neither gives reasons for this assessment nor does he subject his claim to an empirical test. Such a test will be presented below.

3 Method

It was decided to compile a corpus of 1,000 NPs with double *of*-genitives, which divides into 500 spoken and 500 written data points. Distinguishing between spoken and written language provides an opportunity of monitoring the effects of modality and frequency on branching direction (given the empirical observation that recursive *of*-genitives are of unequal frequency in speech and writing). In view of the fact that there is only a binary choice, a sample of 1,000 items was deemed fully sufficient.

The data were culled from the Corpus of Contemporary American English (COCA), a huge electronic data source with a particularly large share of spoken language. The written part of COCA is divided into different genres. Rather than picking examples from different text types, it was decided to rely on a homogeneous data set. The choice fell on academic English, which is certainly a good representative of the written modality.

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

The relevant data were extracted from COCA using the Brigham Young University interface. This interface does not support the simultaneous search for genitive constructions with and without determiners or adjectives. Thus, a total of nine separate search runs had to be performed. One run tested for bare NPs, i.e. NPs with PPs consisting of nothing but a preposition and a noun. Three runs were conducted to identify NPs with one element between the preposition and the noun (more specifically, one intervening element in the first PP, one intervening element in the second PP and one intervening element in each PP). Five runs were launched to find NPs with two elements between the preposition and the noun. (It does not seem necessary to detail them here.) These cover all logical possibilities. The search was terminated at this point, so NPs with three or more words separating the preposition from the noun in the PPs were not included in the sample. It was expected that this omission did not seriously affect the outcome of the study as these complex cases are probably less common than the ones taken into account. All nouns were common nouns. They could be either singular or plural. The search was limited to the year 2008 of COCA.

These nine search runs yielded 903 tokens in the spoken sample and 6,328 tokens in the written sample. The data sets from which these NPs were taken are of similar size (N = 3,930,889 words for the spoken data and N = 4,015,545 words for the written data). This allows us to conclude prior to the data cleansing stage that recursive *of*-genitives are much more often found in written than in spoken English.

The different search runs often identified the same item twice. All these duplications were of course discarded. The search terms automatically generated NPs with more than two *of*-phrases. These were also eliminated. So were all cases in which one *of* was part of a complex preposition (e.g. *by virtue of the connection of the body to a respirator*) or governed by a verb (e.g. *accusing the college of breach of contract*). NPs containing proper nouns were also dropped. In the spoken sample, a number of double *of*-genitives arose as a consequence of using hedges such as *kind of* and *sort of* (e.g. *a combination of sort of curiosity and guilt*). These were also excluded from the data set; so was the single case where an *of* owed its existence to a typographical error (*of* instead of *off*).

This data cleaning procedure brought the 903 tokens in the spoken modality down to 525, of which the first 500 were taken into consideration. Before data cleaning was applied to the written data, the sample was pared down by selecting every seventh hit. Data cleaning reduced the subsample to 621 tokens, which was further reduced by removing every sixth item. From the remaining list of 518 tokens, the first 500 were subjected to analysis.

Branching direction was determined by the author in consultation with two native speakers. Inter-rate agreement was very high (0.99). The few cases of disagreement (N = 4) were either resolved or led to the replacement of a controversial item by the next item on the relevant list.

In view of the complexity of the NPs with recursive *of*-genitives, it is not surprising that the token/type ratio approximates 1. (Two given tokens were treated as instances

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

of the same type when all three nouns were identical.) Thus, the ensuing investigation comes close to a type analysis.

4 Results

Let us begin with a few examples from the corpus. Examples (11) and (12) are from the spoken domain, (13) and (14) from the written domain. Only a minimum of context is necessary for a proper analysis of the data.

(11) General Tony McPeak . . . , former chief of staff of the U.S. Air Force
(12) people of all walks of life
(13) We compared the centers of gravity of public and private courses.
(14) to estimate the prevalence of nonmedical use of prescription and analgesics

While (11) and (13) exemplify left-branching, right-branching is documented in (12) and (14). The designation *chief of staff* is a military rank and hence forms a unit in (11). So does *centers of gravity* in (13). The *gravity of public and private courses* is nonsensical. The unitary status of *walks of life* in (12) is demonstrated by the possibility of replacing it with *professions*. A non-idiomatic NP is shown in (14). Its internal structure can be highlighted by the schema 'the prevalence of X' where 'X' stands for 'nonmedical use of prescription and analgesics'.

The corpus contains a large number of NPs with quantifying expressions because these typically go with the connector *of*. The quantifiers occur either in first or second position, as illustrated in (15) and (16) from the spoken sample, respectively.

(15) after having put a couple of months of works into it
(16) it's a combination of a couple of things

As these two examples show, NPs with quantifiers are inevitably right-branching. This is an immediate consequence of the fact that the quantifier is the head of everything that follows (hence right-branching when it starts off the NP) and completely independent of preceding material (hence right-branching when it takes second position). Owing to their categorical behaviour, NPs with quantifiers were classified separately in the data analysis. However, they were not excluded at the outset because their incidence helps us to assess their role in the 'branching direction game'.

The results of the empirical analysis are reported in table 1.

Table 1 discloses a surprising finding. Contrary to the Garden Path Theory (and in harmony with Plank 2002), right-branching dwarfs left-branching in double *of*-genitives. When NPs with quantifiers are left out of account, 97.4 per cent right-branching contrast with 2.6 per cent left-branching cases. When NPs with quantifiers are taken into account, the disparity is minimally stronger (97.7 as against 2.3 per cent). The inclusion or exclusion of NPs with quantifiers thus has little effect on the branching direction asymmetry. Needless to say, the right-branching preference is highly significant in both conditions (binomial, $p < 0.001$).

Another look at examples (11)–(14) gives us a hint about the forces underlying this bias. The two left-branching cases (11) and (13) contain fixed expressions at the

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

Table 1. *Branching direction in recursive* of*-genitives (N = 1,000) (q. = quantifier)*

Modality	spoken (N = 500)				written (N = 500)			
NPs	with q. (n = 95)		without q. (n = 405)		with q. (n = 32)		without q. (n = 468)	
Branching	left	right	left	right	left	right	left	right
	–	95	15	390	–	32	8	460

beginning of the construction (i.e. *chief of staff* in (11) and *centers of gravity* in (13)). The semantic cohesiveness of these lexical units prevents the association of the second PP with its left-hand neighbour. The only way to leave the fixed expressions intact is to make the second PP their sister constituent. This results in a left-branching structure (see (7)). When a fixed expression occurs at the end of the construction, it cannot help but create a right-branching structure, as exemplified in (12). Obviously, the final PP modifies the adjacent head noun with which it forms a semantic unit.

However, in the absence of fixed expressions, there is a natural bias for right-branching, as documented in (14). When the build-up of recursive *of*-genitives is fully governed by syntactic (rather than partly lexical) principles, the addition of each new PP entails the association with the immediately preceding noun. Such an association is governed by Behaghel's First Law, according to which the linguistic elements which occur next to each other form a conceptual unit. In the case of two PPs, this principle invariably introduces a right-branching bias because the second PP is linked up with the immediately preceding noun, rather than the distant noun which is the head of the entire NP.

The strong right-branching bias in table 1 now emerges as a spin-off function of the fact that syntactic constructions such as double *of*-NPs are typically generated using simple rather than complex lexical units consisting of a noun followed by a (n independent) PP. This follows from the fact that complex lexical units, as defined here, occur far less frequently in the lexicon than simple ones.

The remaining information contained in table 1 can be quickly dealt with. Although there are almost twice as many left-branching cases in the spoken than in the written sample, this difference turns out to be non-significant ($\chi^2(1) = 2.2$, $p > 0.1$). There is thus no statistical basis for the claim that modality (or frequency, for that matter) interacts with branching direction. This result is in harmony with a previous investigation of recursive Saxon genitives (Berg 2011).

By contrast, table 1 reveals a pronounced difference in the frequency of genitive constructions with quantifiers across modalities. These constructions occur three times more often in the spoken than in the written domain. This disparity is mainly brought about by the occurrence of the following three quantifiers: *a bit of*, *a couple of* and in particular *a lot of*. For obvious reasons, these items are not normally found in writing.

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

For one thing, *a lot of* is too colloquial for the written language (Biber *et al.* 1999: 276). For another, all these items are too imprecise in the amount they convey to be of use in academic texts. Because these items ordinarily require an *of*[3] (unlike other indefinite quantifiers such as *many* and *several*), they artificially inflate the subcategory of NPs with quantifiers in the spoken data. As should be evident, this special effect has nothing to do with the issue of branching direction.

Summing up, the foregoing analysis has brought to light a strong right-branching bias in NPs with recursive periphrastic genitives. In keeping with the syntactic nature of these complex NPs, the strength of this bias emanates from a simple principle of syntactic organization. An *of*-phrase that is appended to an immediately preceding head modifies this very noun rather than any other. In the event of recursive *of*-phrases, this iconic principle imposes right-branching on the full NP. The empirical observation that this constraint is not a categorical one can be ascribed to the fact that the organization of these complex NPs is not completely syntactic. Certain *of*-phrases form a lexical unit with the head they modify and accordingly create a left-branching structure when these lexical units occur at the beginning of the NP. In this view, the rate of left-branching reflects the rate of idiomatic expressions consisting of NOM and PP (in this particular syntactic position). In the following, the nature of the relationship between branching direction and recursion will be probed more deeply.

5 Theoretical discussion

In this section we will cast our net wide to compare a number of recursive patterns and their branching direction preferences. The default situation would seem to be a consistent relationship between recursion and branching direction in that all recursive structures, no matter what type they are, exhibit either left-branching or right-branching. Recourse to the Saxon genitives, as discussed in section 2, makes it plain that this simple assumption has to be rejected. Whereas periphrastic genitives display a pronounced penchant for right-branching, Saxon genitives gravitate towards left-branching. Thus, the relationship between recursion and branching direction is a complex one. Our task is, then, to unravel the factors leading to left-branching in one case, but to right-branching in another.

We begin by introducing a few more empirical data. Nominal compounding is also recursive in English. As was mentioned in the introductory section, tripartite compounds allow for both left- and right-branching. It has been confirmed by several studies that these complex compounds evince a left-branching bias (e.g. Warren 1978; Krott *et al.* 2004; Kösling & Plag 2009). Berg (2011) showed that this left-branching bias is equally strong in nominal compounds and Saxon genitives.[4]

[3] In the present context, it may safely be ignored that *of* tends to be deleted in *a couple of* in some varieties of the spoken language.

[4] Note in parentheses that a left-branching proclivity also emerges in four-noun compounds (Berg 2006).

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

A further case of recursion is represented by relative clauses. These structures are widely regarded as prototypical instances of recursion in the linguistic literature (e.g. Palmer 1971: 151). A well-known example is *the cat that chased the mouse that stole the cheese that* . . . However, it is worth pointing out that although recursive relative clauses are perfectly acceptable, they have a very low frequency of occurrence in everyday language. In any event, it is quite obvious that NPs with double relative clauses are invariably right-branching (e.g. Kuno 1973: 8). The explanation offered above for NPs with periphrastic genitives tells us why. Left-branching in NPs with periphrastic genitives was argued to require idiomatic expressions. However, the typical formal makeup of idioms is not an antecedent followed by a relative clause.[5] The connection between these two parts is too loose to allow them to be integrated into a single lexical unit. Relative clauses are syntactic objects *par excellence*. No wonder, then, that like NPs with recursive periphrastic genitives, NPs with recursive relative clauses lean towards right-branching, only more extremely so.

Let us return to NPs with double Saxon genitives. Berg (2011) demonstrated that these form a heterogeneous set. A major factor influencing the internal structure of NPs with recursive Saxon genitives is the animacy of the three nouns, especially that of the first possessor. If all three nouns are animate, the NP is invariably left-branching, as in (17); if, however, at least one of the three nouns is inanimate, both branching directions are observed, as exemplified in (9) and (10) above.

(17) she was also his sister's husband's sister

The left-branching nature of (17) derives from the fact that the head noun *sister* is preceded by the complex modifier *sister's husband*. Henceforth, cases like (17) for example will be treated as a separate subcategory and referred to as 'personal Saxon genitives'.

We now have surveyed four types of recursive structures, each with a distinct branching direction bias. These four types may be ordered on the scale depicted in figure 1 according to their branching direction preference.

Occupying the extreme ends of the scale we find personal Saxon genitives and relative clauses. The former always branch on the left, the latter always on the right.[6] On the face of it, the reasons for this difference are all too obvious. In a sense, personal Saxon genitives and relative clauses are mirror images of each other. The head of a personal NP with two Saxon genitives always occurs on the right (as is characteristic of morphological heads in English; see Williams 1981) and hence is premodified. By contrast, the head of an NP with recursive relative clauses is always left-peripheral and hence postmodified. It is well known that there is a relationship between the order of heads and modifiers and branching direction (e.g. Dryer 1992). Heads are lexical while

[5] Such cases are notably absent from the corpus of fixed expressions culled and investigated by Moon (1998).

[6] The categorical behaviour of these two classes assigns them to the first type introduced in the opening section (i.e. branching direction determined by syntax only). However, as will be argued below, the boundary between the different categories is less strict than one might think.

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press



Figure 1. Branching direction bias in different recursive patterns

modifiers are phrasal in nature. By themselves, heads cannot be phrasal because, if they were, they would have to be internally complex, that is, consist of a head and a modifier. Because the phrasal nature is the prerequisite for branching – lexical units cannot branch – the position of the phrasal element determines branching direction. Thus, NPs with relative clauses are right-branching, but NPs with personal Saxon genitives left-branching.

NPs with double *of*-genitives are akin to NPs with double relative clauses, both empirically and conceptually. As shown before, they display a strong propensity for right-branching because the head is modified on the right. However, there is a crucial difference between the two types of NP in terms of the nature of the internal boundaries. By definition, NPs with relative clauses have an internal clausal boundary, which separates the antecedent from the relative clause. In contrast, NPs with periphrastic genitives evince phrasal, though no clausal boundaries. The syntactic 'depth' of the boundary has an important implication. The 'deeper' the boundary, the more strongly lexical chunking is discouraged. Therefore, lexical chunking may not occur in NPs with relative clauses, but may occur in NPs with periphrastic genitives. As lexical chunking is the only means to establish cohesion between a head noun and a following *of*-genitive, it is the only way to overcome the right-branching bias imposed by postmodification. The possibility, or otherwise, of lexical chunking accounts for the minor empirical difference in branching direction between NPs with *of*-genitives and those with relative clauses.

This line of argument may be usefully extended to nominal compounds. When compounds and NPs with *of*-genitives are compared in terms of the depth-of-boundary hypothesis, we find shallower boundaries in the former than in the latter type. While NPs with periphrastic genitives show phrasal boundaries (see above), compounds by their very nature lack these.[7] With only morphological boundaries separating its constituents, a tripartite compound is virtually unrestricted in its susceptibility to lexical chunking effects. Thus, any chunk can appear in any position of a compound. To be specific, a three-noun compound may in principle accommodate a two-part chunk as readily in positions 1 and 2 as in positions 2 and 3. All other things being equal, we should accordingly expect the same rate of left- and right-branching in tripartite compounds. However, all other things are not equal. The garden path effect gives an edge to left-branching in compounds (Berg 2012). The same effect contributes to the left-branching bias in NPs with recursive Saxon genitives taken as an undifferentiated group (not represented in figure 1).

[7] A curious exception to this generalization is the so-called phrasal compounds (e.g. Lieber 1988).

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

We now have reviewed three major factors that enter into a general account of branching direction, viz. the order of heads and modifiers in conjunction with their lexical and phrasal nature, lexical chunking and the garden path effect. The first is a syntactic, the second a lexical and the third a processing principle. The syntactic factor is a categorical constraint in that the expansion of a head can only be on the left or on the right. Moreover, it is more general than the other factors. Lexical chunking is a local constraint in that it involves individual combinations of lexemes occurring in certain syntactic or morphological positions. It is neutral with respect to branching direction. The processing factor is a relatively weak constraint which facilitates left-branching, but does not prohibit right-branching.

These three factors, their individual strengths as well as their interaction, give rise to the empirical patterns in recursive structures. The syntactic factor can be observed in isolation in NPs with recursive relative clauses. Both lexical chunking and the garden path effect are inoperative here because the clausal boundary prevents chunking and because there is sufficient syntactic information to prevent structural misassignments. In contrast, NPs with lesser syntacticity are open to lexical effects. These effects play a marginal role in NPs with double *of*-genitives (owing to their relatively high degree of syntacticity), but a more central role in NPs with double Saxon genitives (owing to their low degree of syntacticity). The garden path effect can only play a role when there is room for structural misassignment. The higher the number and reliability of available phrase-building elements, the lower the risk of a garden path. Hence, the greater the syntacticity of an NP, the weaker the garden path effect. Putting this another way, its scope is limited to lexical and mildly syntactic NPs such as compounds and NPs with recursive Saxon genitives.

It remains for us to explain within the present model the empirical difference between NPs with personal Saxon genitives (as defined above) and other NPs with double Saxon genitives. As will be recalled, the former are exclusively left-branching whereas the latter allow for both branching directions. Let the following be noted at the outset. The fact that right-branching is not attested in the sample of NPs with personal Saxon genitives does not imply that it is banned once and for all. It does not take much imagination to conjure up an *X's X* construction (with *X* standing for an animate noun) which is idiomatic and therefore quite cohesive (e.g. *bitch's son* as an unattested alternative to *son of a bitch*).[8] If this artificial unit was premodified by a possessor (i.e. animate noun plus possessive marker), we would end up with a right-branching structure. This reasoning argues against a categorical difference between NPs with personal Saxon genitives and the others.

The absence of idiomatization in personal Saxon genitives is the reason for their consistently left-branching structure. In this type of NP, the genitive marker signals kinship or other forms of social relationship such as *my brother's girlfriend* and *her*

[8] A not too dissimilar type of idiomatization is actually attested with inanimate possessions such as *writer's block* and *driver's licence*. I can see no reason why such cases should be categorically ruled out with animate possessions.

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

husband's commander, with the former occurring more frequently in the data than the latter. Critically, these relationships do not differ in strength, or else the language does not code any potential differences in strength. For instance, there is no coding difference between *(my) father's teacher* and *(my) teacher's children*. So when these two genitive constructions are combined into the NP *my father's teacher's children*, a symmetry is created between the two links. This symmetry prevents a closer association between the second and the third noun and therefore stands in the way of right-branching. As previously explained, the categorical left-branching bias in personal Saxon genitives emanates from their leftward expansion in conjunction with the garden path effect.

By contrast, NPs with one or more inanimate nouns need not always show this symmetrical relationship. Saxon genitives with inanimate nouns may be subject to mild forms of idiomatization (see footnote 8) and hence bear the potential of right-branching (provided the idiomatic expression occurs in the right margin of the NP).

In addition to the factors discussed above, there is a further processing principle influencing branching direction both in qualitative and quantitative terms. This is Hawkins's (1994) approach to assessing processing complexity in terms of constituent recognition domains.[9] The basic idea is that the shorter the recognition domain, the lower the number of nodes that have to be concurrently active and, hence, the lower the processing cost. Crucially, language structure is assumed to be sensitive to processing complexity, so the less costly a given pattern is, the more strongly it is preferred. In the language comprehension process, constituents are created on the basis of constituent-building units, which are usually the heads of constituents. The size of a recognition domain is defined as the distance from one constituent builder to the next. Processing complexity in turn is defined as the ratio of immediate constituents to the number of words in the recognition domain. A simple example may illustrate the calculation procedure. Compare (18) to (19).

(18) She put down the receiver.

(19) She put the receiver down.

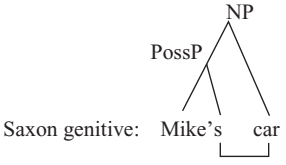
In terms of processing complexity, (18) is to be preferred to (19). The recognition domain for the phrasal verb is higher in the former than in the latter construction. Hawkins's ratio for (18) is perfect because there are three constituents, the first two of which consisting of a single word, so $3/3 = 1$. In (19), by contrast, the recognition domain spans four words, so we obtain a lower value (given the same number of constituents): $3/4 = 0.75$.

Hawkins's theory provides a straightforward explanation for the fact that Saxon genitives occur to the left, and *of*-genitives to the right of their head nouns. Their constituent recognition domains are shorter than if it was the other way around. Consider (20):

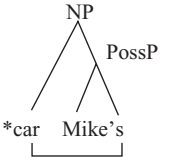
[9] I am very grateful to an anonymous *ELL* referee for making this apposite observation.

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

(20)

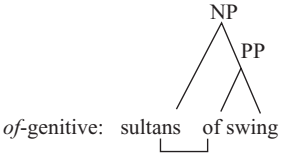


(20′)

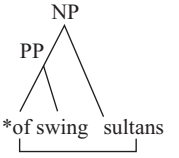


The possessive marker is a constituent builder as it creates the possessive phrase *Mike's*. In the left-branching structure (20), the minor and the major head (i.e. the possessive marker and the possession) are adjacent, so the constituent recognition domain is optimal. In the artificially constructed right-branching structure (20′), however, the constituent-to-morpheme ratio is lower at $2/3 = 0.67$. This difference nicely explains the left-branching nature of Saxon genitives in English. Now consider (21):

(21)



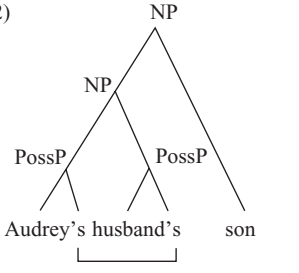
(21′)



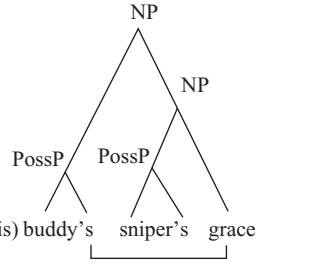
As in the case of the Saxon genitive, the constituent builders are adjacent in the well-formed example (21) but non-adjacent in the ungrammatical example (21′). The constituent recognition domain is perfect in (21), as the number of constituents equals the number of words (i.e. $2/2 = 1$). As before, the value for the unattested order is appreciably lower, i.e. $2/3 = 0.67$. This difference explains the right-branching nature of English *of*-genitives.

Hawkins's theory accounts not only for the grammaticality of divergent head-modifier ordering in the two types of genitives, but also for the relative frequency of left- and right-branching decisions in the recursive patterns at the heart of this study. Let us compare the recognition domains in Saxon genitives, as in (22) and (23).

(22)



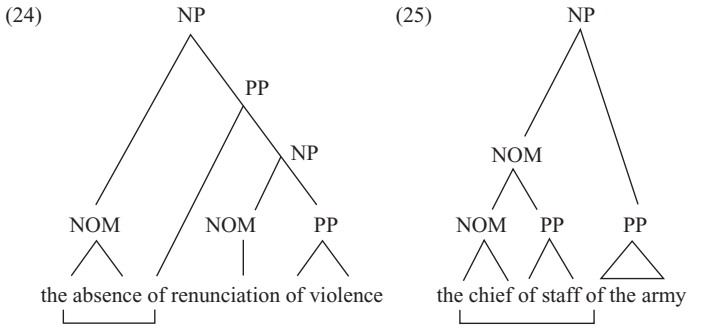
(23)



The distance between the first possessive marker and the head of its sister constituent is smaller in (22) than in (23). The constituent-to-morpheme ratio amounts to $2/3 = 0.67$ in the former case but $2/4 = 0.5$ in the latter. The shorter recognition domain gives an edge to left-branching over right-branching in recursive Saxon genitives, as empirically confirmed in Berg (2011).

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

Why we observe the opposite branching direction preference in recursive *of*-genitives is visualized in (24) and (25).



It takes less time to identify the beginning of the sister constituent (or the end of the current constituent, for that matter) in the right-branching structure (24) than in the left-branching structure (25). The constituent-to-word ratio is $2/3 = 0.67$ in (24) but only $2/5 = 0.4$ in (25). Note that the two NPs in (24) and (25) are not structurally identical. There is a determiner in the second PP in (25) but none in (24). However, this disparity is inconsequential as the determiner follows the preposition constructing the PP and is in any case irrelevant in the process of constructing the sister constituent. Once again, the ease-of-recognition hypothesis proves eminently successful. It correctly predicts the predominance of right-branching in recursive *of*-genitives.

It is worthwhile to add that Hawkins's framework might also account for the existence of both categorical and probabilistic patterns in language. As Hawkins (1994) points out, the larger the difference in constituent-to-word ratios between two given structures, the higher the likelihood of one being licit and the other illicit. Inversely, the smaller this difference, the higher the probability of both being grammatical, though differentially frequent. A comparison of the differences in percentage of the various ratios reveals that these differences tend to be more pronounced in the case of the categorical (non-recursive) than in the probabilistic (recursive) patterns. The former show a difference of 33 per cent, the latter a difference of 17 and 27 per cent. These individual data points are no more than suggestive. While the difference between 33 and 27 per cent is certainly not impressive, 33 per cent is twice as much as 17 per cent. Thus, it may be tentatively argued that Hawkins's processing theory is powerful enough to elucidate the nature of both categorical and probabilistic decisions in language. From this psycholinguistic perspective, the difference between grammatical and ungrammatical patterns on the one hand and frequent and infrequent patterns on the other emerges as a gradual one. In fact, the two patterns find a natural place on the same continuum.

To conclude, it has been argued that branching patterns in recursive *of*-constructions are brought about by a conspiracy of several factors of unequal strength. The syntactic factor imposes constraints on lexical chunking. Through the absence of syntactic marking, it also determines where misassignments may occur. Given the contrast between syntax and lexicon, lexical structures are more prone to garden path effects

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

than syntactic ones. The factors in question differ not only in terms of strength, but also in terms of the branching direction that they introduce. Whereas the processing factor favours left-branching, the lexical factor is inherently neutral with respect to branching direction. The syntactic factor influences branching direction by deciding between leftward and rightward expansion. It seems that these three factors in conjunction with Hawkins's theory of processing efficiency suffice to account for the empirical patterns.

6 Outlook

Let us finish by pointing out a few diverse directions which future work might take. An obvious extension of the present study would be to look at even longer structures, that is, NPs with three *of*-genitives. In fact, the electronic search for double *of*-genitives also unearthed a certain number of triple *of*-genitives (even in the spoken corpus). Not surprisingly, the great majority of these were right-branching, although there were also a few exceptions. Another natural extension would be to examine the branching direction of recursive structures in languages other than English. It may be of interest to ascertain whether the same factors as in English are at work and how their strength varies from language to language. The nature of these factors strongly suggests that they have some cross-linguistic validity. Finally, a quite unexpected spin-off of the present work is the light it sheds on the syntactic constraints on idiomatization. This is a rather neglected topic in idiom research. The present study leads us to predict that the proportion of idiomatic to compositional expressions is a function of their syntactic cohesiveness. The deeper the syntactic boundary within a given unit, the smaller the likelihood of idiomatization. Furthermore, idiomatization is a matter of degree. Therefore, it may additionally be predicted that the degree of idiomatization is all the higher, the shallower the internal syntactic boundary inside phrases. These predictions await lexicon-wide testing.

Author's address:
Department of English
University of Hamburg
Von-Melle-Park 6
20146 Hamburg
Germany
thomas_berg@uni-hamburg.de

References

Berg, Thomas. 2003. Right-branching in English derivational morphology. *English Language and Linguistics* 7, 279–307.

Berg, Thomas. 2006. The internal structure of four-noun compounds in English and German. *Corpus Linguistics and Linguistic Theory* 2, 197–231.

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press

Berg, Thomas. 2011. Recursion introduces a left-branching bias (where possible). *Linguistics* 49, 977–90.
Berg, Thomas. 2012. *Structure in language: A dynamic perspective*. New York: Routledge.
Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.
Bolinger, Dwight & Donald A. Sears. 1981. *Aspects of language*. New York: Harcourt Brace Jovanovich.
Burton-Roberts, Noel. 1986. *Analysing sentences: An introduction to English syntax*. London: Longman.
Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68, 81–138.
Frazier, Lyn & Keith Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14, 178–210.
Haspelmath, Martin. 2002. *Understanding morphology*. London: Arnold.
Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
Kösling, Kristina & Ingo Plag. 2009. Does branching direction determine prominence assignment? An empirical investigation of triconstituent compounds in English. *Corpus Linguistics and Linguistic Theory* 5, 201–39.
Krott, Andrea, Gary Libben, Gonia Jarema, Wolfgang U. Dressler, Robert Schreuder & Harald Baayen. 2004. Probability in the grammar of German and Dutch: Interfixation in triconstituent compounds. *Language and Speech* 47, 83–106.
Kuno, Susumu. 1973. *The structure of the Japanese language*. Cambridge, MA: MIT Press.
Lieber, Rochelle. 1988. Phrasal compounds in English and the morphology–syntax interface. In Diane Brentari, Gary Larson & Lynn MacLeod (eds.), *Papers from the 24th annual regional meeting of the Chicago Linguistic Society. Part Two: Parasession on agreement in grammatical theory*, 202–22. Chicago: Chicago Linguistic Society.
Moon, Rosamund. 1998. *Fixed expressions and idioms in English: A corpus-based approach*. Oxford: Clarendon Press.
Palmer, Frank. 1971. *Grammar*. Harmondsworth: Penguin.
Plank, Frans. 1985. How disgrace-ful. *Acta Linguistica Hafniensia* 19, 64–80.
Plank, Frans. 2002. Grouping in multiple attribution: Advantage Albanian. In Wolfram Bublitz, Manfred von Roncador & Heinz Vater (eds.), *Philology, typology and language structure*, 157–70. Frankfurt am Main: Peter Lang.
Rayner, Keith, Marcia Carlson & Lyn Frazier. 1983. The interaction of syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior* 22, 358–74.
Warren, Beatrice. 1978. *Semantic patterns of noun–noun compounds*. Göteborg: Acta Universitatis Gothoburgensis.
Williams, Edwin. 1981. On the notions 'lexically related' and 'head of a word'. *Linguistic Inquiry* 12, 245–74.

https://doi.org/10.1017/S1360674312000160 Published online by Cambridge University Press