8

Constructions, complexity, and word order variation

PETER W. CULICOVER

8.1 Introduction

This chapter is concerned with the possibility of accounting for word order and word order variation in terms of complexity[1]. I propose that it is useful to consider word order variation in terms of competing constructions, where other things being equal, the less complex construction is preferred by speakers. This view of variation presupposes that we have a way of measuring complexity. I suggest that both formal complexity and processing complexity play a role in driving change and variation. Formal complexity has to do with how much idiosyncrasy and irregularity has to be

---

[1] This chapter is an adaptation and extension of a number of topics discussed in Culicover 2013. It was made possible by the generous support of the Alexander von Humboldt Stiftung. I am grateful to the participants in the Conference on Formal Linguistics and the Measurement of Grammatical Complexity, University of Washington, March 23rd and 24th, 2012, and audiences at the Ohio State University, Humboldt University in Berlin, the University of Tübingen, the University of Wuppertal, and UC Davis for useful questions and comments. I thank also Erhard Hinrichs, Marga Reis, Andrzej Nowak, Susanne Winkler, Ray Jackendoff, Guido Seiler, Jack Hawkins and Martin Salzmann for their support and suggestions. I owe a particular debt to Laurel Preston and Fritz Newmeyer for their detailed and constructive editorial feedback. Any errors are my responsibility.

Since space is limited, I take the opportunity to cite here the work that has inspired much of my thinking about these issues.

- Markedness and optimality (Chomsky 1964, 1965)
- Processing complexity and dependency (Gibson 1998, 2000; Hawkins 1994, 2004; Haider 1995, 1997, 2000, 2003)
- Empirical work on verbal clusters and word order in Continental West Germanic [CWG] (Wurmbrand 2004, 2005; Haegeman and van Riemsdijk 1986; Zwart 1995, 1996; Bader and Schmid 2009a, 2009b, 2009c; Schmid 2005; Schmid and Vogel 2004; Vogel 2003, 2009; Bies 1996; Sapp 2005, 2006, 2007.)
- Factors bearing on word order (Lötscher 1978; Wasow 1997, 2002, 2009; Hawkins 1982, 1994, 1999, 2001, 2004, 2006)
- Linearization and structure (Curry 1961; Dowty 1996; Kathol 2000)
- Social dynamics and epidemiology (Nowak and Latané 1994; Seiler 2008; Enfield 2008)
- Constructional approaches to grammar (Fillmore et al. 1988, Kay and Fillmore 1999)

expressed by a grammar, while processing complexity concerns the computational resources required by a language user to map between syntactic forms and conceptual structures. Focusing on English infinitival relative clauses and Continental West Germanic verb clusters, I suggest several ways in which complexity may be measured and how such complexity may contribute to language change and variation. I consider how complexity may actually arise in the course of change, and why it may persist even in the face of pressures to reduce it.

I originally became interested in the issue of constructional variation when I was working on the evolution of English *do*-support (Culicover 2008). One of the functions of *do*-support is to preserve the adjacency of the main verb and its complement, even in questions where there is inversion, as in (1).

(1)   Did any of the students [pass the exam]?

In contrast, in a V2 language that lacks *do*-support, the verb often becomes separated from its complements, as in (2).

(2)   *Old English*
        ða   wendon   hi   me   heora   bec   to.
        then   turned   they   me   their   backs   to
        'Then they turned their backs to me.'
        [Allen 1980: 287]

It is at least plausible that because *do*-support permits reduction of the dependency distance, it is less complex in some sense than the V2 alternative. In fact, Hawkins 2004 has proposed that the grammars of languages are organized in such a way as to minimize dependency distance.

However, assuming that in general grammars value simplicity and efficiency, explaining *do*-support in terms of the reduction of dependency distance raises an obvious puzzle: why don't all languages have *do*-support and eliminate the V2 alternative? That is, why are there languages like German and Dutch?

In the case of *do*-support it is plausible that there are in fact multiple dimensions of grammatical complexity that play a role in privileging each construction over the other. On the one hand, there is an advantage to introducing a construction such as *do*-support, which reduces dependency distance. On the other hand, there is a competing advantage to a V2 language, which preserves the form of the verb regardless of whether the verb is adjacent to its complements (as in (1)), or separated from its complements (as in (2)). The advantage afforded by the V2 language is arguably that the thematic structure governed by the verb is more readily identified when the main verb is inverted than when a dummy modal such as *do* is inverted. On this scenario, pressure to reduce complexity on one dimension may conflict directly with pressure to reduce complexity on another dimension. Languages or dialects may differ in that one pursues reduction of complexity on one dimension, while another pursues reduction of complexity on another dimension.

But then the question arises: why does one complexity factor win out in some languages, while another factor wins out in others? My suggestion is that the choice of which factor wins out when there are competing factors is not strictly a linguistic matter, but reflects social factors as well.

The structure of this chapter is as follows. §8.2 addresses the question: How do we move from intuitions about complexity on various dimensions to some concrete measurement of complexity? I distinguish two types of complexity, formal complexity and processing complexity, and discuss how they might be measured.

Section 8.3 focuses on a well-documented instance of word order variation, the Continental West Germanic (CWG) verb clusters. The simplest examples of these clusters consist of sequences of two verbs, in the order V_1-V_2 or V_2-V_1, where V_2 is the head of the complement of V_1 in a standard syntactic representation as in (3).

(3) $[_{VP} V_1 [V_2 \ldots]]$

There are also three-verb clusters, which show all possible orderings of V_1, V_2, and V_3, V_3 being the most deeply embedded verb; some of these clusters are very well-attested, while others are very rare. The data shows that in many language varieties[2] it is not possible to say that all two- or three-verb clusters show a particular order. Which order is possible depends on the lexical subclasses of V_1, V_2, and V_3 that participate in each particular ordering. This sensitivity of order to lexical subclass suggests that word order is specified by CONSTRUCTIONS, that is, form–meaning correspondences. I discuss how to formulate such constructions in §8.4.

Proceeding from the assumption that alternative orders may emerge as a way of reducing complexity on different dimensions, in §8.5 I propose two dimensions of complexity that may account for the observed variation in CWG verb clusters.

The role of social factors in explaining the distribution of properties is taken up in §8.6. Given the properties and distributions of the CWG verb clusters, I offer the following scenario: Social factors, such as network topology, geography, and different frequency distributions over speakers may favor one constructional alternative over another in different geographical areas, leading to variation. Contact leads to spread of properties, resulting in mixed varieties. I illustrate these points with a computational simulation. Section 8.7 concludes with a summary.

8.2 Measuring grammatical complexity

Two of the most prominent notions of grammatical complexity in the literature are what I refer to here as formal complexity and processing complexity.[3] Formal

---

[2] 'Language variety' is a cover term used in sociolinguistics for languages, dialects, registers, and so on.
[3] For other terminology and related discussion, see the papers in Miestamo (2008), as well as the chapters by Newmeyer & Preston and Moran & Blasi in this volume.

complexity has to do with the degree of generality of a grammatical description, in the spirit of markedness theory (Chomsky 1964, 1965; Battistella 1996). The intuition underlying this notion of complexity is that more general phenomena are simpler and more 'natural'. Irregularity, exceptionality and idiosyncrasy contribute to greater complexity, and, in principle at least, this complexity can be measured in terms of the number of terms, statements and length of statements in a description of the phenomena using some standard formal vocabulary.[4]

Processing complexity, on the other hand, has to do with the computational resources that are required by language users to map between a string of words and an interpretation. It is assumed in syntactic theory that processing complexity is not represented in the grammar per se (cf. e.g. Chomsky and Miller 1963). It has been proposed that differences in processing complexity may explain frequency differences in constructions with the same function, and, in the limit, the non-occurrence of a particular construction (Hawkins 1994, 2004).

For concreteness, I give one example of each type of complexity. I show in §8.2.1 how reduction in formal complexity may be appealed to in order to explain some instances of language change. In §8.2.2 I turn to processing complexity, and in particular, its role in determining the frequency of competing constructions in the linguistic experience of learners. In the discussion of CWG verb clusters to follow, I argue that since the formal complexity of the competing constructions is the same, the variation should be understood in terms of processing complexity.

8.2.1 *Formal complexity: an example*

Formal complexity can be illustrated by the English infinitival relative. Fillers in infinitival relatives must be a PP (4b), rather than an NP (4a).

(4)   a. *the man who to talk to __
      b. the man to whom to talk __

In this way, they differ from infinitival questions, where the filler may be an NP.

(5)   a. I wonder who to talk to __.
      b. I wonder to whom to talk __.

The grammaticality of (5a) suggests that the ungrammaticality of (4a) is an idiosyncrasy that must simply be stipulated in the grammar of English (Sag 1997; Culicover 2011).[5]

---

[4] See Chater and Vitányi (2007) for discussion of the relationship between simplicity, description length, and language acquisition.

[5] For discussion of a number of other syntactic idiosyncrasies, see Culicover 1999, 2013. The impossibility of (4a) has been recognized in generative grammar for quite some time (cf. Emonds 1976: 192–195), but has not always been viewed as a gap in an otherwise regular pattern.

Comp. by: PG0844   Stage : Proof   ChapterID: 0002130152   Date:28/4/14   Time:18:29:33
Filepath:d:/womat-filecopy/0002130152.3D152

OUP UNCORRECTED PROOF – FIRST PROOF, 28/4/2014, SPi

152   *Constructions, complexity, and word order variation*

The presence of idiosyncrasy is somewhat puzzling on the view that languages seek to reduce formal complexity. One might think that all exceptions would be eroded away over time. English infinitival relatives offer an intriguing insight into how complexity as exemplified by (4a) may actually arise as a language changes in the direction of greater generality. In Old English (OE), infinitival relatives lacked relative pronouns entirely (Kemenade 1987: 151). They used either zero or the equivalent of *that*. When the relativized phrase was a constituent of PP, the clause was a zero-relative and the preposition was stranded, as illustrated in (6).

(6) Drihten, Þu   Þe   gecure Þæt   fæt   [PP on *t*]   to eardienne
     Lord,   you   yourself chose   that vessel   in *t*   to live
                                               [Blick[6] 157/ Kemenade 1987: 151]

But finite relatives had relative pronouns. The OE tensed relatives did not permit P-stranding with a clause-initial *wh*-; only clause-initial pied-piped relative PP was possible (Kemenade 1987: 152–53). It is therefore plausible to assume that infinitival *wh*-relatives are an innovation. The innovation was in the direction of increasing uniformity, by extending *wh*-relatives from the tensed to the infinitival cases. The scenario is illustrated in Fig. 8.1. Stage 1 is the extension of *wh*-relatives to infinitivals, and Stage 2 is the extension of infinitival relatives to PPs.

The generalization began by extending relatives from zero-tensed to zero-infinitival. The next stage, still manifested by present-day English, extended infinitival relatives to clause-initial PP, leaving out clause-initial NP. The resulting form is an idiosyncrasy when interpreted with respect to tensed relatives, which allow all three types, as seen in Figure 8.1. In other words, the pressure to reduce formal complexity in the general domain of relative clauses resulted in an increase in formal complexity in the more restricted domain of infinitival relatives, marked by the stipulation in the grammar that the overt relative in an infinitival relative must be a PP.

Relatives	zero	PP	NP
Tensed			
Infinitival	(Stage 1)	(Stage 2)	

FIGURE **8.1** Evolution of relative clause types in English.

---

[6] 'Blick' refers to the Blickling Homilies, dated around the end of the 10th century (Westlake 1908).

A likely possibility, given *wh*-NP in initial position in infinitival questions (i.e. *wonder who to talk to*), is that the last cell will be filled in over time, and in fact isolated cases of this case can be found, as shown in (7):

(7) a. Where do I find *the person who to* talk to about the quest?
      http://answers.yahoo.com/question/index?qid=20100727153550AAPs5ed

   b. In this case, I'll refer to the radio/club DJ as being *the person who to* target.
      http://independentmusicstartup.com/440/how-to-get-a-tastemaker-to-take-your-music-to-the-next-level/

   c. All requests for aid should include:…3. The name and identity of the requesting person or the *person who to* contact upon arrival.
      http://www.jdcap.org/SiteCollectionDocuments/EmergencyPlanExternal.pdf

   d. That I needed the name of the Mortuary and number and *the person who to* speak to to verify if it was true…
      http://www.prisontalk.com/forums/archive/index.php/t-488639.html

My native speaker judgment is that these examples are quite acceptable, in contrast with (4a), but only if the head noun is *person* (and not, for example, *man*). It is possible that they reflect the earliest stages in the elimination of the idiosyncrasy that constitutes formal complexity, namely, that the relative marker in an infinitival relative must be PP. Of course, it cannot be known at this point whether the change will generalize or simply die out.

8.2.2 *Processing complexity*

The second type of complexity is relevant to the processing of syntactic structures and assigning them interpretations. This type of complexity may be observed in terms of measures such as eye-tracking, self-paced reading, and reaction times (Gibson 1998, 2000; Levy 2005, 2008; Demberg and Keller 2008). For example, it is known that subject relatives (*the doctor that consulted the nurse*) require more processing (as measured by reading time) than object relatives (*the doctor that the nurse consulted*), although both types of relatives are fully grammatical. Self-embedding contributes further to processing complexity, so that sentences like (8c) are virtually impossible to comprehend.

(8) a. The doctor visited the patient.
   b. The doctor that the nurse consulted visited the patient.
   c. The doctor that the nurse that the hospital hired consulted visited the patient.

Hale (2001) proposes that the greater difficulty of processing configurations such as object relatives and self-embedding correlates with the frequency of occurrence of such configurations in corpora, and thus the probability of the configuration. The lower the probability of a particular configuration, the higher the 'surprisal' at

Comp. by: PG0844   Stage : Proof   ChapterID: 0002130152   Date:28/4/14   Time:18:29:33
Filepath:d:/womat-filecopy/0002130152.3D154

OUP UNCORRECTED PROOF – FIRST PROOF, 28/4/2014, SPi

the point at which the processor is forced to switch from the expected parse to an unexpected parse (Hale 2001, 2003; Levy 2005, 2008). In extreme cases, the sequence of words up to a particular point in the sentence points strongly to one continuation, but the actual continuation is different, resulting in a garden-path effect (Frazier 1987).

The question then arises, what makes one construction less frequent than another? In some cases, at least, the answer may be that the less frequent construction is more complex in terms of production. Complexity has been attributed to the maintenance of representations in memory (Kluender 1998), the maintenance of reference and the cost of the operations that build structure (Gibson 2000), and the cost of backtracking and repair (Frazier 1987). On the assumption that speakers avoid the more complex in favor of the less complex (Hawkins 1994, 2004), other things being equal, the frequency of more complex constructions in the experience of learners will be lower. Hence there is a loop that links processing complexity for the speaker to frequency to processing complexity for the hearer.

In sum, there is a plausible case to be made for relating processing complexity, corpus frequency and acceptability, although the precise details of the relationship remain to be worked out. In the next section I summarize the data from CWG verb clusters, as a step towards outlining a scenario for understanding variation in terms of processing complexity.

8.3 CWG verb clusters

Verb clusters (exemplified in English by *will sing, has sung, try to find,* etc.) are useful for exploring complexity and variation for two reasons. First, they show a substantial amount of variation among forms that have the same interpretation, and can therefore be compared directly in terms of formal and processing complexity. Second, there is a wealth of empirical data about the distribution of various types of verb clusters over the Continental West Germanic area, so that we can seek correlations among various types of clusters. This section summarizes the basic types of clusters and suggests that they should be analyzed as individuated constructions. The next section addresses how a constructional analysis of verb clusters might be formulated.

Verb clusters are sequences of modals (MOD), *have-* and *be*-type auxiliaries (AUX), and various control verbs (V). Typical two-verb clusters are given in (9), using Standard German forms for the sake of illustration. The verb order is shown in parentheses, where 1 is the highest verb (i.e. the one closest to the root) in a standard syntactic analysis, 2 the next highest, and so on (cf. (3)). So in (9a), for example, 1 is *kann* and 2 is *singen.*

(9)   Maria glaubt, daß
      Maria believes that

    a. sie die Arie singen kann. (2–1)
       she the aria sing can
       '…she can sing the aria.'
    b. sie die Arie kann singen. (1–2)

The order 2–1 of (9a) is found in Standard German; the 1–2 order of (9b) is found in a number of Dutch dialects, Swiss German dialects and other CWG varieties. I suggest in §8.5 that 1–2 is a response to pressure to order scope taking elements, such as AUX and MOD, before their complements, while 2–1 is a response to pressure to position heads as close as possible to their dependents.

The 3-verb clusters found in CWG are shown in (10), again using standard forms. 'Rare' means that the particular ordering is found only in a few varieties, and is highly lexically restricted.

(10)   a. sie Peter die Arie singen hören wird. (3–2–1)
       she Peter the aria sing hear will
       '…she will hear Peter sing the aria'
    b. sie Peter die Arie hören singen wird (2–3–1) [rare]
    c. sie Peter die Arie wird hören singen (1–2–3)
    d. sie Peter die Arie wird singen hören (1–3–2)
    e. sie Peter die Arie singen wird hören (3–1–2)
    f. sie Peter die Arie hören wird singen (2–1–3) [rare]

An important property of CWG verb clusters is that some of them are very frequent, others are very rare, and there are apparent correlations between two-verb and three-verb clusters. That is, certain three-verb clusters occur only in varieties with 1–2 in two-verb clusters, while others occur only in varieties with 2–1. Many varieties appear to permit more than one order, both in two- and three-verb clusters.

Figure 8.2 shows the distribution of two-verb clusters in nine different varieties (data is from Sapp 2011: 108; the headings are renamed for consistency).

Sapp highlights the fact that for each variety and for each subtype of cluster, there are different preferences for word orders (not shown in Figure 8.2). For example, he cites Patocka 1997 as observing that in the Austrian dialects, the passive with *werden*-V.PAST. PRT is uniformly 2–1, while *haben*-V.PAST.PRT permits both orders in some cases.

Along similar lines, Dubenion-Smith 2010:112 found the judgments in Figure 8.3 in an analysis of spoken West Central German[7] from the *Datenbank Gesprochenes*

---

[7]   The area bounded roughly by Karlsruhe to the South, Darmstadt to the East, Kassel to the North, and Aachen to the West.

Dialect	MOD V	AUX V
Standard German	2–1	2–1
German & Austrian dialects (Wurmbrand)	2–1	2–1
S and W Austria	1–2 (2–1)	1–2 2–1
N Austria	2–1	2–1
E Austria	2–1	1–2 2–1
Bavarian	2–1	2–1 (1–2)
Swabian	2–1	2–1 (1–2)
Alsatian	2–1 (1–2)	2–1
Swiss	1–2 (2–1)	2–1 (1–2)

FIGURE 8.2 Word order in two-verb clusters (Sapp 2011: 108).

Deutsch compiled between 1955 and 1970. Each cell shows the number of tokens and percentage of the total number of tokens of each ordering for verb clusters containing verbs of particular classes, e.g. past participle V and AUX=*haben* 'have', *sein* 'be', *hätten* 'have.SUBJUNCTIVE', *wären* 'be.SUBJUNCTIVE', or *werden* 'Passive be', in the first row, V MOD in the second row, and so on. The data show that both 1–2 and 2–1 are possible for many two-verb clusters, and that different combinations of verb classes occur with different frequencies. (I have changed some of Dubenion-Smith's notations for the sake of uniformity.)

Dubenion-Smith's more detailed figures, not reproduced here, show that for V AUX, V MOD, V *tun* and $V_2 V_1$, both of the orders 1–2 and 2–1 are attested, with a strong preference for 2–1. Dubenion-Smith takes the useful step of breaking down the data into specific subtypes of two-verb clusters, which allows the variability to show through clearly. While the totals show that both orders are possible, the

Syntagm (group)	2–1	1–2	Total tokens
V AUX (perfect, subjunctive, passive)	887 (92.4%)	73 (7.6%)	960 (72.4%)
V MOD	227 (74.9%)	76 (25.1%)	303 (22.9%)
V *tun* 'do'	27 (96.4%)	1 (3.6%)	28 (2.1%)
V_2V_1	13 (59.1%)	9 (40.9%)	22 (1.7%)
V *kriegen* 'get'	8 (100%)	0 (0%)	8 (0.6%)
V *lassen* 'let/make'	4 (100%)	0 (0%)	4 (0.3%)
MOD AUX	1 (100%)	0 (0%)	1 (0.1%)
Totals	1,167 (88%)	159 (12%)	1,326 (100%)

FIGURE 8.3 Two-verb clusters in spoken West Central German subordinate clauses (Dubenion-Smith 2010: 112).

breakdown shows that the 1–2 order is more common for MOD V (as a percentage of the total cases) than it is for the other subtypes.

Similar variability is found in three-verb clusters. Consider the data in Fig. 8.4 (from Schmid 2005) showing possible three-verb clusters in Zürich German, broken down by the type of verb that occupies the V2 slot. The data show that which three-verb clusters are possible in this variety depends on the verb types of V_1 and V_2. (The verb types are exemplified by *lassen* 'let/make' (Causative), *müssen* 'must' (Modal), *sehen* 'see' (Perception Verb), *helfen* 'help' (Benefactive), *bleiben* 'stay' (Durative), *beginnen* 'begin' (Inchoative), and *versuchen* 'try' (Control Verb).)

V_2	$AUX_1\ V_2[\text{PAST.PRT}]\ V_3$	$AUX_1\ V_2[\text{INF}]\ V_3$ (IPP)	$FUT_1\ V_2\ V_3$
Causative	*	3–2–1, 1–2–3, 1–3–2	3–2–1, 1–2–3, 1–3–2
Modal	*	?3–2–1, 1–2–3, 1–3–2	?3–2–1, 1–2–3, 1–3–2
Perception verb	3–2–1, ?1–2–3, 2–1–3	?2–3–1, 1–2–3	3–2–1, 1–2–3, 1–3–2
Benefactive	3–2–1, 2–3–1, 1–2–3, 1–3–2, 2–1–3	2–3–1, 1–2–3, 2–1–3	3–2–1, 2–3–1, 1–2–3, 1–3–2
Durative	3–2–1	*	3–2–1, 1–3–2
Inchoative	2–3–1, 2–1–3	*	2–3–1, 1–2–3, 2–1–3
Control verb	3–2–1, 1–2–3, 2–1–3	*	3–2–1, 1–2–3, 1–3–2, 2–1–3

FIGURE **8.4** Overview of verb order patterns in Zürich German (adapted from Schmid 2005: 76).

According to Schmid, "Zürich German shows the largest variation of verbal order patterns of all languages [. . .] With the exception of order 3–1–2 (only possible with a special stress pattern), all logically possible patterns are confirmed by my inform-ant."[8] But what her summary also shows is that different orders are attested with different verb classes.

Data such as these show that it is not possible to say that two- and three-verb clusters in a particular language variety have a particular ordering. In some language varieties, particular orderings occur only with particular classes of lexical items. In some language varieties, more than one ordering occurs, but one order occurs significantly more frequently than the other.

The variety of attested orderings raises a number of questions: How do we account for the possible orderings in each variety? Specifically, is there a derivational account that explains the observed orderings in terms of an underlying 'canonical' order? Or must the various orderings be treated as distinct, but related, constructions? Why are

---

[8] However, according to another native speaker informant (Martin Salzmann, p.c.), not all of these orders are actually active in this dialect. It is entirely plausible, although of course difficult to confirm, that the judgments of Schmid's consultant were influenced by exposure to standard German and to other dialects of Swiss German.

Comp. by: PG0844   Stage : Proof   ChapterID: 0002130152   Date:28/4/14   Time:18:29:34
Filepath:d:/womat-filecopy/0002130152.3D158

OUP UNCORRECTED PROOF – FIRST PROOF, 28/4/2014, SPi

some orderings more frequent than others? Does either formal or processing com-
plexity have anything to do with these phenomena? If the rare orderings are more
complex in some sense than the more common ones, then why have they not been
completely supplanted by the less complex orderings? How are clusters properly
integrated into grammatical descriptions in terms of syntax and semantics? That is,
what is the relationship between the structure of a verb cluster and its interpretation?
I propose some answers to these questions in the following sections.

8.4 Constructions

There are three properties of verb clusters that have to be accounted for. First,
clusters with the same elements but different orders have the same interpretation.
Second, all of the possible orders of clusters composed of the same elements occur in
at least one language variety. And third, the clusters 1–2–3 and 3–2–1 are significantly
more frequent than the others.

In this section I propose a constructional analysis, in which the verb clusters are
individuated unheaded phrases. That is, they are constituents, members of a sui
generis category, call it VerbCluster (VC).[9] Individual variants of a word order (e.g.
AUX_1-V_2 vs. V_2-AUX_1) are distinct but related constructions, that is, they are form-
meaning correspondences (in the sense of Culicover and Jackendoff 2005). On this
view, a construction consists of a pairing of a syntactic configuration, and a concep-
tual structure. Crucially, the related constructions express the same correspondence
between form (syntactic structure) and meaning (conceptual structure).

In order to show the explanatory power of the constructional account, I first
outline the more traditional derivational approach, in which a verb cluster with a
particular word order is derived from some canonical underlying structure by
movement and adjunction (Wurmbrand 2004, 2005).[10] The synonymy of clusters
with different orders is accounted for by deriving them from the same underlying
structure, which corresponds to a particular interpretation.

Consider, for example, *das Buch lesen₂ kann₁* 'can read the book', with the
structure in (11a), and the alternative *das Buch kann₁ lesen₂* with the structure in

---

[9] For extensive arguments that syntactic theory should allow for sui generis constructions that exist in a
language alongside more general constructions, see my book Culicover (1999). In Culicover (2013), I discuss
the nature of the formal relationship between the more specific and the more general constructions in a
language.

[10] I do not pursue such an approach in detail here, in part because of Wurmbrand's demonstration that
there is no evidence that favors one candidate underlying order from another, in part because it is not clear
how to account for the variability of the sort noted in the preceding section in terms of movement and
adjunction, and in part because it does not offer a natural way to account for the relative rarity of some
orderings versus others.

(11b), derived by movement from (11a). The corresponding conceptual structure for both expressions is (11c), where **READ**(AGENT:X, THEME:BOOK) corresponds to *das Buch lesen* and **ABLE** corresponds to *kann*.

(11) a. $[_{VP} [_{VP} [_{NP} \text{das Buch}] \text{lesen}_2] \text{kann}_1]$
      b. $[_{VP} [_{VP} [_{NP} \text{das Buch}] t_2] \text{kann}_1+\text{lesen}_2]$
      c. **ABLE**(**READ**(AGENT:X,THEME:BOOK)

Since the function **ABLE** takes as its argument the relation **READ**(AGENT:X, THEME: BOOK), it is customary to say that the modal scopes over its complement in both (11a) and (11b).

Different orders are accounted for on such a derivational account by stipulating that certain heads adjoin to the left or to the right of higher heads, along the lines of (11b). More complex derivations are required for three-verb clusters; a few are illustrated in (12), assuming underlying 3–2–1.

(12) a. $[_{VP} [_{VP} [_{VP} \ldots V_3] V_2] V_1] \Rightarrow [_{VP} [_{VP} [_{VP} \ldots t_3] V_2+V_3] V_1]$ (2–3–1)
      b. $[_{VP} [_{VP} [_{VP} \ldots V_3] V_2] V_1] \Rightarrow [_{VP} [_{VP} [_{VP} \ldots t_3] V_2+V_3] V_1]$
          $\Rightarrow [_{VP} [_{VP} [_{VP} \ldots t_3] t_{2+3}] V_1+V_2+V_3]$ (1–2–3)   [the + should not be subscripted]
      c. $[_{VP} [_{VP} [_{VP} \ldots V_3] V_2] V_1] \Rightarrow [_{VP} [_{VP} [_{VP} \ldots V_3] t_2] V_1+V_2]$ (3–1–2)
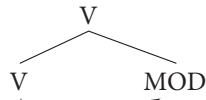
As far as I know there is no natural mechanism intrinsic to the grammar in such an account to explain why some orders are very frequent and others are not. One might surmise that the number of adjunctions would correspond in some sense to the complexity or markedness of a particular cluster. The application of two adjunctions shown in (12b) would then predict that 1–2–3 is very rare or possibly non-existent. However, the 1–2–3 and 3–2–1 clusters are the most common.[11]

Let us consider now the constructional account. The problem of how to interpret the clusters is dealt with by unification, along the lines proposed by Bouma and van Noord (1998). In unification, the features of the constituents that form a phrase are assigned to the phrase itself. The semantics do not require hierarchical branching structure in the syntax, because the scope relation is established in the correspondence with conceptual structure: the representation corresponding to one verb takes the representation headed by the next as its argument. The selectional properties of the verbs are unified, as shown in (13) for a two- and in (14) for a three-verb cluster with two modals (e.g. *lesen*$_3$ *können*$_2$ *wird*$_1$ '(s/he) will be able to read').
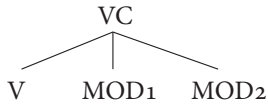
---

[11] If the underlying order were 1–2–3, then the incorrect prediction would be that 3–2–1 is rare.

Comp. by: PG0844   Stage : Proof   ChapterID: 0002130152   Date:28/4/14   Time:18:29:34
Filepath:d:/womat-filecopy/0002130152.3D160

OUP UNCORRECTED PROOF – FIRST PROOF, 28/4/2014, SPi

160   *Constructions, complexity, and word order variation*

(13)

SYNTAX

CS $\mathbf{F}[\varphi_M](V'[\varphi_V]) \Rightarrow \mathbf{F}(V')[\varphi_M \cup \varphi_V]$

(14)

SYNTAX

CS $\mathbf{F1}[\varphi_{M1}](\mathbf{F2}\varphi_{M2}]([V'[\varphi_V])) \Rightarrow \mathbf{F1}(\mathbf{F2}(V'))[\varphi_{M1} \cup \varphi_{M2} \cup \varphi_V]$

So, following (13), the interpretation of [*lesen₂ kann₁*] will be (15), where **ABLE** corresponds to **F**, φ_M represents the selectional properties of *kann*, **READ** corresponds to **V′**, and φ_V represents the selectional properties of **READ**, here, the argument structure (AGENT:X,THEME:Y).

(15)   **ABLE**(**READ**)[AGENT:X,THEME:Y]

Similarly, the interpretation of (16a) will be (16b).

(16)   a. lesen₃   können₂   wird₁
           read.INF   can.INF   will
           'will be able to read'
       b. **FUT**(**ABLE**(**READ**))[AGENT:X,THEME:Y]

In addition to (13) and (14) there exist constructions for each of the other possible orders, each of which has the same interpretation. For example, the construction for MOD₁-V₂, e.g. *kann₁ lesen₂* 'can read', will show the same correspondence with **ABLE** (**READ**) as *lesen₂ kann₁*, and the construction for MOD₁-MOD₂-V₃, e.g. *wird₁ können₂ lesen₃* will show the same correspondence with **FUT**(**ABLE**(**READ**)) as *lesen₃ können₂ wird₁*.

Consider now the fact that that the constructional approach presupposes that there are as many individuated constructions as there are distinct word orders with the same interpretation. From the perspective of description this is a virtue, since in fact all orders are attested. Moreover, the existence of all of these constructions does not render the constructional account overly complex, if we incorporate into our syntactic theory the further assumption that all orderings of the daughters of a phrase in a construction are available in a grammar, if at least one ordering of the daughters

of that phrase are. In this respect I am drawing on an intuition that has been implemented in a variety of ways in the theoretical literature, most importantly in optimality theory (see for example Schmid and Vogel 2004), as well as in varieties of categorial grammar (Muskens 2007).

Given this assumption, the existence of a construction where V_2-V_1 corresponds to a specific interpretation immediately entails that V_1-V_2 is possible with the same interpretation. I assume, however, that whether a given construction is actually used or accepted by native speakers depends in part on the frequency with which they encounter instances of that construction, in part on the formal complexity of the construction with respect to the grammar as a whole (as discussed in §8.2.1), and in part on the processing complexity of the construction with respect to the available alternatives. The assumption that alternative orders of phrasal constituents are in principle generally available is critical, because it allows for an explanation of the spontaneous emergence in a variety of a construction as well as the growth of a competing construction under the influence of contact. Of course, not all orders are equally likely to occur, in part for reasons of complexity, as I discuss below.

The principled availability of all orderings of a construction also permits an explanation of why the 1–2–3 and 3–2–1 three-verb clusters are more frequent than the others, and why 2–3–1 and 2–1–3 are quite rare. 1–2–3 is optimal with respect to one of the complexity biases discussed in the next section, and 3–2–1 is optimal with respect to the other. As noted above, this asymmetry is not accounted for in a derivational account. It is in fact stipulated in OT accounts of CWG verb clusters (Bader and Schmid 2009a,b,c; Schmid and Vogel 2004), but not explained in terms of complexity. The constructional account I propose accounts for the relative frequency of the various orders in terms of complexity and the dynamics of the social network, as discussed in the next two sections.

8.5 The role of complexity biases in accounting for change and variation

Section 3 summarized data showing that the ordering in verb clusters is variable, across and within varieties, and even within speakers. I suggest in what follows that 1–2 is a response to pressure to order scope taking elements, such as AUX and MOD, before their complements, while 2–1 is a response to pressure to position heads as close as possible to their dependents. Why multiple orders can be simultaneously maintained is a matter that I take up in §8.6.

To get an insight into the biases that affect word order in verb clusters, consider the following quotation from Haider. Haider (2003: 28) argues that because of 'parser unfriendliness' there must be a constraint that requires branching nodes to follow the heads, thereby ruling out the left branching structure. Instead, there must be a cluster in V-final structures. (See also Dubenion-Smith 2010 for a similar idea.)

"In a right branching structure, the parser can unambiguously identify the top-most node of the projection after encountering the first element of the projection, V1. [The left branching structure], however, is not parser-friendly. The parser would have to guess how many brackets there might be, because their number—or in other words, the depth of embedding of the left-most element—depends on the number of verbs to come. General top-down information on the possible structure of a VP will not help guessing, because the number of auxiliaries is not context dependent." (Haider 2003: 8)

A slightly different way to put Haider's observation is that there is nothing formally wrong with 3–2–1 order, per se, but 1–2–3 order is computationally less complex with respect to a particular aspect of processing. The tense or modal operator expressed by the first verb must take scope over the VP that is its argument, as discussed in §8.4. Hence, it appears that something along the lines of the following is correct (see Kroch 1974, for example).

(17)   *Scope-Order principle*
       The preferred scope ordering of operators corresponds to the left-to-right ordering of the phrases in the surface structure of the sentence.

This principle suggests the following scope bias.

(18)   *Scope bias*
       Alignment of scope with linear order facilitates one aspect of the computation of scope in the CS representation.

With this in mind, consider what steps are required in processing a 2–1 order, exemplified in (19).

(19)   ... daß sie das Buch lesen$_2$ will$_1$.
        that she the book to-read wants
        '... that she wants to read the book.'

For simplicity, let us suppose that there is a certain cost to processing a verb that selects a VP complement that precedes it, and that the costs associated with each such verb are equal; let this measure be c, which I will call the 'scope bias'. This cost occurs because when the preceding verb is encountered, it is not known precisely what the following verb will be, or even whether there will be such a verb in some cases, as Haider suggests. Since we don't know exactly how the parsing proceeds, I assume that the cost of each basic operation is also c, to simplify the exposition and to make it simpler to compare alternative ways of resolving parsing conflicts and uncertainties.

At the point at which the word *lesen* 'read' is processed, the partial representation is (20).

(20)   **READ**(PRO, BOOK)

A potential complication is that *sie* is singular and the singular form of *lesen* is *liest,* so *sie* is not the subject of *lesen* but of some other verb that follows *lesen*. However, *sie* has the same form as *Sie* 'they, you (polite)', and the infinitival form and the 3.PL inflected form are the same. So at the point of *lesen*, there is no way to predict that there will be another verb, as Haider points out. Either we hold *lesen* in memory as an ambiguous form, or we decide that it agrees with *Sie*.[12]

In either case there is additional processing at *will* 'want', which is ambiguously 1|3. SG. If the processor refuses to make a commitment regarding *lesen*, then it has to hold the form in memory, a cost of **c**, by assumption. If the processor decides that *lesen* was 3.PL, then it has to change the interpretation from *Sie* 'they, you (polite)' to *sie* 'she/her', which also costs **c**. In either case, the result is (21), with a marginal processing cost of **c**, also by assumption.

(21)   **WANT**(PRO[3.FEMALE]$^\alpha$, **READ**(α,BOOK))

Suppose now that the order is 1–2. In this case there is no marginal cost to processing the first verb in the cluster, because of its position in the sequence. Processing ... *daß sie das Buch* is the same in both cases. Then the processor gets to *will*, which shows that the subject is the 3.SG *sie* and not 3.PL *Sie,* which takes a different form of the verb (*wöllen*).

(22)   **WANT**(PRO[3.FEMALE]$^\alpha$, **F**(α,BOOK))

And when the processor encounters *lesen*, all that is required is to fill in the value of **F** as **READ** and link the object thematic role to BOOK.

Processing the order 2–1 is thus apparently more complex than processing the order 1–2 by **c**. So, one might ask, why aren't all of the dialects 1–2(–3) (not to mention all languages)? The answer, as suggested earlier, is that there is also a dependency bias that favors 2–1 and 3–2–1. I state it here as follows, paraphrasing Hawkins's (1994, 2004) formulation in terms of syntactic domains—the syntactic domain of a head H and its dependent D is the smallest part of the tree that dominates both H and D.

[Should be Hawkins']

(23)   *Dependency bias*
         The preferred ordering of a head and its dependents is the order that permits
         the minimal syntactic domain that contains them.

As discussed at length by Hawkins (1994, 2004), languages tend to favor linearizations in which the subcategorized arguments of a head and closely related adjuncts

---

[12] If the first verb is a past participle, such as *gelesen* 'read.PAST.PRT', it is very likely that the following verb is a form of the auxiliary verb *haben* 'to have', but the exact form of this verb cannot be predicted. It could be infinitival *haben* or finite (present or past tense), and in the latter case, it is inflected for person and number.

are as close to the head as possible, thereby reducing the domain in which the dependency can be computed by the processor. Hawkins's reasoning is that the computational cost of constructing the CS representation, which is correlated with memory load, is greater when the heads and dependents are more distant from one another. A similar measure of complexity is argued for by Gibson (1998, 2000) and Grodner and Gibson (2005).

Consider again the processing cost of 1–2 versus 2–1. Let us assume that holding an argument or an adjunct in memory until its head is encountered incurs a cost of c'; this is the dependency bias. In the 2–1 variant of our example (19), there is no cost, since the direct object *das Buch* is adjacent to the verb *lesen*. But in the 1–2 variant, the verb *will* intervenes:

(24)   . . . daß sie das Buch | will lesen.

The processor may hypothesize at the point that *das Buch* is encountered that it is an argument of some verb, probably a direct object, and so corresponds to $F(\alpha,\text{BOOK})$. But just as holding $READ(\alpha,\text{BOOK})$ until *will* is processed incurs a cost when the order is 2–1, so does holding $F(\alpha,\text{BOOK})$ until *lesen* is processed in the 1–2 cluster. Hence with respect to processing the dependency, processing 1–2 is more costly than processing 2–1 by c'. And when we deal with more complex VPs, the processing costs are correspondingly greater for both orders.

Order 1–2 is thus favored by the scope bias, and 2–1 by the dependency bias. Competing biases do not necessarily cancel one another. The constructions that they apply to are both active in the population of speakers, and in principle one or the other can acquire an advantage through an accidental property of the topology of the network. That is, a critical mass of speakers can shift to one order for non-linguistic reasons, making it the only order or the preferred order.[13]

Crucially, these two biases are in competition only in languages with VP-final clusters. In a language like English, where V is initial in VP, the 1–2(–3) cluster is optimal with respect to both the scope bias and the dependency bias. In languages with VP-final clusters, however, alternative orders may be seen as responding to one or the other of these biases. (25) illustrates, where X represents the complements of the main verb.

---

[13] What I am suggesting here is similar to, but different from, the notion of 'competing motivations', which as I understand it typically attributes different outcomes of language change to competition between formal and functional pressures, and between various functional pressures. In the case under discussion, I am proposing that different contributors to processing complexity are in competition and produce different outcomes in different social contexts. For discussion see Newmeyer (2005) and references cited there.

(25) X 3–2–1   ⇒ X **1–3–2**   [scope bias]
      X 1–2–3   ⇒ X **2–1–3**   [dependency bias]
      X 1–2–3   ⇒ X **2–3–1**   [weak dependency bias, assumes
                              unitary 2–3; cf. Haider 2003]
      X 1–2–3   ⇒ X **3–1–2**   [dependency bias]

All of these orders are attested, as shown below. 1–3–2 is the order found in the Infinitivo Pro Participio construction AUX-V-MOD, where MOD has the infinitival rather than the participial inflection, as in (26).

(26)   . . . daß er das Buch hat lesen können/*gekonnt
        that he the book has read.INF can.INF can.PST.PRT
        '. . . that he was able to read the book.'

The order 2–1–3 is found in Zürich German, as shown in (27).

(27) a. dass i en gsee₂ ha₁ schaffe₃
          that I him seen have.1S work.INF
          'that I saw him work'
          (M. Salzmann, p.c.)

      b. Wo s aagfange₂ hat₁ rägne₃, . . . [ZüGe.]
          when it begin.INF has rain.INF
          'It began to rain.'
          (Lötscher 1978)

Notice that in the case of (27a), at least, 2–1–3 puts the verb closer to its overt argument, satisfying the dependency bias. A similar case can be made for (27b), assuming that *s* 'it' is a thematic argument of *aagfange* 'begin.INF'. 2–1–3 apparently occurs only with verbs such as *anfangen* 'begin' and *sehen* 'see', which again supports the analysis of verb clusters as individuated constructions. That is, it is not correct to say that Zürich German allows the order 2–1–3 in three-verb clusters. Rather, the 2–1–3 order occurs with particular verbs, an idiosyncrasy that is straightforwardly expressed by the construction in (29).

(28)
    SYNTAX   VC

              V1   'have'   V2

[please boldface
marked expressions] CS $\boxed{\text{PAST}}[\varphi_{\text{AUX}}](\boxed{V1'}[\varphi_{V1}](\boxed{V2'}[\varphi_{V2}]))) \Rightarrow \boxed{\text{PAST}}(\boxed{V1'}(\boxed{V2'}))[\varphi_{\text{AUX}} \cup \varphi_{V1} \cup \varphi_{V2}]$

        where V1={'see', 'begin',...}

2–3–1 clusters with a range of verbs are well-documented in Afrikaans (Biberauer
n.d); e.g.,

(29) a. ...dat hy die medisyne kon   drink   het   [modal]
          that he the medicine could.INF drink.INF have
          '...that he could drink the medicine'

      b. ...dat hy hom die medisyne maak/ laat drink het   [causative]
          that he him the medicine make.INF let.INF drink.INF have
          '...that he made/let him drink the medicine'

      c. ...dat hy haar hoor   roep   het   [perception]
          that he her hear.INF call.INF have
          '...that he heard her call'

      d. ...dat ek haar die bokse help   dra   het   [benefactive]
          that I her the boxes help.INF carry.INF have
          '...that I helped her carry the boxes'

      e. ...dat die mense bly staan het   [durative]
          that the people remain.INF stand.INF have
          '...that the people remained standing'

      f. ...dat dit ophou   reën   het   [inchoative]
          that it stop.INF rain.INF have
          '...that it has stopped raining'

      g. ...dat hy probeer voorgee het   [control]
          that he try.INF pretend.INF have
          '...that he tried to pretend'

      h. ...dat hy die boek gaan lees het   [motion]
          that he the book go.INF buy.INF have
          '...that he went to buy the book'

      i. ...dat hy die boek loop (en) koop het   [linking]
          that he the book walk.INF and buy.INF have
          '...that he went and bought the book'

      j. ...dat hy die boek sit en lees het   [linking]
          that he the book sit and read have
          '...that he was sitting and reading the book'

According to Biberauer (n.d.), this 2–3–1 pattern alternates with 1–2–3, and is possible
only when the finite verb is a form of *hebben* 'have'. The idiosyncrasy of this pattern
is captured naturally as a construction that is similar to but different in detail from
(29)—in Afrikaans the 2–3–1 pattern has become a distinct construction, one that is
restricted to a particular finite V_1 but is very free with regards to the V_2 in the cluster.
Finally, 3–1–2 is the standard order in Swiss German dialects that are uniformly 2–1, a
point that is discussed further in the next section.

These data suggest that the verb clusters are constructional, and not derived by movement and adjunction. Constructions are designed to accommodate any degree of idiosyncrasy in a uniform descriptive framework. While it is possible in principle to express variation in derivational terms, using parameters, the degree of idiosyncrasy that is found in verb clusters (and I have presented only a portion of the data) would require parameters of such lexical specificity that the theory of parameters would be rendered vacuous. The parameters would effectively be a way to encode constructions in the grammar without calling them constructions. (For an extended argument along these lines, see Culicover 1999.)

Furthermore, there is evidence that a lexically restricted construction may generalize as it spreads geographically. Generalization is seen in Seiler's (2004) discussion of Swiss German verb clusters. These clusters display a number of intriguing patterns and correlations. For example, V_2-V_1 (using Standard German forms: *lesen₂ lassen₁* 'make read') is not found unless V_2-MOD$_1$ (*lesen₂ können₁* 'can read') is found, and V_1-MOD$_2$ is not found unless V_2-AUX$_1$ (*gelesen₂ haben₁* 'have read') is found. At the same time, 3–1–2 (*lesen₃ kann₁ haben₂* 'read can have') occurs only if the order 2–1 is fully general (across AUX, MOD and main V), suggesting a generalization of 1–2 > 2–1 to the three-verb case ('place main verb first in the sequence'). In this regard, Seiler writes (p. 13): "First, I have shown that the ordering of elements in western dialects is strictly ascending (1–2–3), but as the more we move eastwards the more the tendency for ascending ordering weakens. Second, the ordering of elements is sensitive to the category of the head. Auxiliaries tend most to be set at the right edge of the cluster. **This tendency is much weaker with modal verbs and almost absent with lexical verbs as heads of a cluster.**" (Emphasis mine—PWC)

This type of lexical idiosyncrasy is quite extensive in verb clusters. One way to interpret the Swiss data is that the 1–2(–3) pattern was originally widespread. Then the 2–1 order began as an innovation stimulated by the dependency bias, where 1 is AUX; the construction was subsequently generalized to MOD and finally V. On this view, the order 3–1–2 is a final generalization of 2–1, extending the change motivated by the dependency bias.[14]

This sensitivity to category and even to individual items is what we would expect if individual verb clusters are distinct constructions. The apparent generalization from one category to another (e.g. the generalization from auxiliaries to modals to lexical verbs) is also fully consistent with an account in which there is pressure to reduce formal complexity as marked by lexical idiosyncrasy, as discussed in §8.2.1.

A pattern similar to that found in Swiss German is seen in Heerlen Dutch (Cornips 2009). In this variety, 2–1 is preferred for V1=AUX and 1–2 is strongly preferred for

---

[14] Whether such changes actually occurred must unfortunately remain speculative for now. The survey of the Swiss German dialects from which Seiler's analysis is drawn dates from 2000. As far as I know there are no comparable surveys from earlier times that would provide the suitable diachronic perspective.

V1=MOD. One interpretation is that there is a change going on from 1–2 to 2–1 which is applying first to V1=AUX. This interpretation is consistent with that of Cornips (p. 212), who argues for basic 1–2 on the basis of the correlations between order preferences of individual speakers.

In West Flemish, 2–1 is required for V1=AUX and 1–2 is required for V1=MOD (Haegeman 1994). Again, this pattern suggests a shift from 1–2 to 2–1 beginning (and ending) with AUX-V.

If a variety permits 1–2 or 2–1 for a particular pair of verbs, then by iteration, it should also permit 1–(2–3) and (3–2)–1 respectively for longer combinations of these verbs. Hence we would not expect to find 3–2–1 in the absence of 2–1, or 1–2–3 in the absence of 1–2. If a variety mixes 1–2 and 2–1, as West Flemish and Afrikaans do for V1=AUX and MOD, we expect 2–3–1 where V1=AUX, V2=MOD, since V follows MOD and AUX is maximally final in the sequence. Wurmbrand (2005) identifies West Flemish as 2–3–1 in the IPP construction[15]; e.g. (32) from Haegeman (1994).

(30)   ..da Valère nie nor us   will$_2$-en   kom$_3$-en   eet$_1$
        that Valery not to   house want-INF   come-INF   has
        '..that Valery did not want to come home.'

I have already suggested that 3–1–2 may be motivated by the dependency bias applying in a 1–2–3 variety. The coexistence of MOD$_1$-MOD$_2$-V$_3$ and V$_3$-MOD$_2$-AUX$_1$ in West Flemish raises the possibility that the scope bias for two modals is stronger than the dependency bias, but that the dependency bias wins when there is only one modal.

In summary, if a variety permits 1–2 or 2–1 for a particular pair of verbs, then we would expect that it permits 1–(2–3) and (3–2)-1 respectively as constructional generalizations. Crucially, we do not expect to find 3–2–1 in the absence of 2–1, or 1–2–3 in the absence of 1–2 (and we don't find it, as far as I can tell).

I have proposed in this section that verb clusters in which the order of elements corresponds to their scope in the semantic interpretation are the consequence of a processing bias that favors such a correspondence, while verbs clusters in which the order has the elements with wider scope following are the consequence of a

---

[15]   The Infinitivus Pro Participio construction is characterized by the fact that V$_1$ is an AUX, typically a form of 'have', and V$_2$ is MOD in the infinitival form, not the participial form. In Standard German, the IPP cluster is 1–3–2 and not the otherwise typical 3–2–1. E.g.,

(i)   …daß Peter das Buch hat lesen können
       …that Peter the book has read.INF can.INF
and not

(ii)   a. *daß Peter das Buch lesen gekönnt hat
           that Peter the book read.INF can.PAST.PRT has
        b. *daß Peter das Buch lesen können hat
           that Peter the book read.INF can.inf has (Schmid 2005)

processing bias that favors positioning heads close to their dependents. The 1–2–3 clusters uniformly reflect the scope bias, while 3–2–1 clusters uniformly reflect the dependency bias. Clusters with mixed orders, such as 2–1–3 and 2–3–1, may be analyzed as reflecting both biases. Such mixed orders appear to be lexically restricted, a fact that has a natural characterization in constructional terms. Finally, it appears that lexical restrictions on a verb cluster may be relaxed or lost completely as the construction is generalized under the pressure to reduce formal complexity.

8.6 A computational simulation

Consider now the fact that in spite of differences in complexity, many different verb clusters are attested. Some clusters are far from optimal with respect to one or another bias. A natural question to ask is why complexity is maintained in the face of pressures to reduce it? Using a computational simulation, I show how a bias can lead to the reduction in the frequency of use of a construction in a population and even eliminate a construction entirely. I also show, using this simulation, that it is possible for complexity to be maintained if the social network has a suitable configuration. I hasten to note at the outset that a simulation is not a proof, but simply a potentially useful way of visualizing how an observed state of affairs may have come about.

The simulation involves a set of agents arranged in an array. These agents influence one another through contact.[16] Since we are simulating language, the agents are understood to be speakers. In the illustrations I give here, the array is 50*50. That is, there are 2500 speakers in this social network. I assume an initial state of affairs prior to contact as shown in Figure 8.5. There are three binary features which together define eight possible languages in this population of speakers. The three feature values are initially distributed randomly over the population, as shown in upper and lower right quadrants, and the lower left quadrant in Figure 8.5. One value of a feature is represented by a black square, the other by a white square. Each agent has one of the two feature values for each feature. The composite of the feature values for the three binary features is shown in the upper left quadrant. Each agent in this quadrant has one of eight possible feature combinations, represented by grad-ations on the gray scale from white to black. The simulation assumes that at the outset there is an even distribution of the possible feature combinations across the population, so that all eight languages are attested and have roughly the same share of the population.

At each step of the simulation, agents influence one another with respect to each feature of variation. The value of a feature of one agent will change if more of its

---

[16] This illustration is adapted from chapter 6 of Culicover (2013), from which Figs. 5–13 are taken. The computational model is due originally to Nowak and Latané (1994) and is described at some length in Culicover and Nowak 2003.

Comp. by: PG0844   Stage : Proof   ChapterID: 0002130152   Date:28/4/14   Time:18:29:35
Filepath:d:/womat-filecopy/0002130152.3D170

OUP UNCORRECTED PROOF – FIRST PROOF, 28/4/2014, SPi

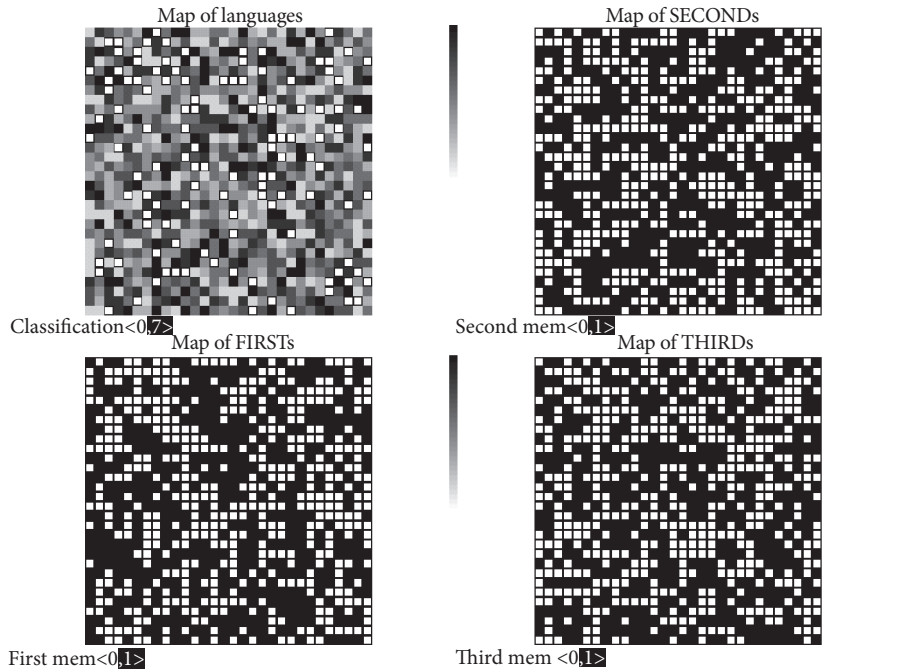170   *Constructions, complexity, and word order variation*

FIGURE 8.5 Feature values at initial state.

interaction partners have the other value. In this very simple simulation, all of the agents have equal influence on the others with whom they are in contact, and all of the feature values and combinations of feature values are of equal strength.[17] So, for example, if agent A is interacting with two agents, and one has the value Black and one has White, then the influence of each on A is equal. For convenience, we set the parameters of the simulation model so that A does not change from its current value unless confronted with a greater number of the alternative value. If A has White and interacts with two agents, one of which has Black and the other of which has White, A stays with White. But if the two agents both have Black, then A changes to Black. Figure 8.6 illustrates.

Figure 8.6 illustrates the distribution of languages and features after 69 steps of a simulation in which the agents in Figure 8.7 influence one another and change as a consequence of their interactions.

---

[17] It is possible in more complex simulations to weight the effect of distance between interacting agents, the number of agents that interact with one another, the strength of individual agents, individual resistance to change, and other parameters of the simulation model.
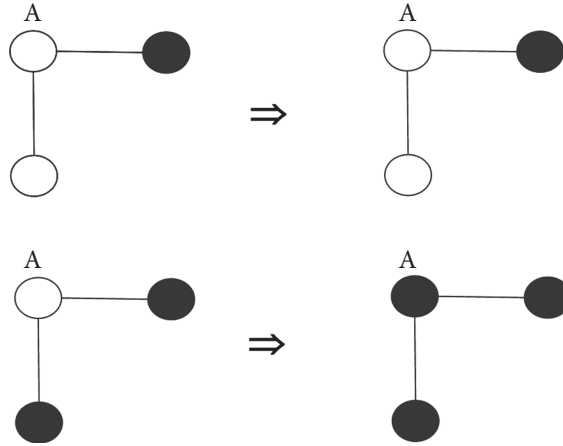
Comp. by: PG0844   Stage : Proof   ChapterID: 0002130152   Date:28/4/14   Time:18:29:35
Filepath:d:/womat-filecopy/0002130152.3D171

OUP UNCORRECTED PROOF – FIRST PROOF, 28/4/2014, SPi

FIGURE 8.6 Interactions of neighboring agents in a network. The state of the network prior to interaction is shown to the left of the arrow, and the state of the network after interaction is shown to the right of the arrow. All interactions are computed and executed simultaneously, so it is possible that an agent may influence one agent but itself be influenced to change by other agents.
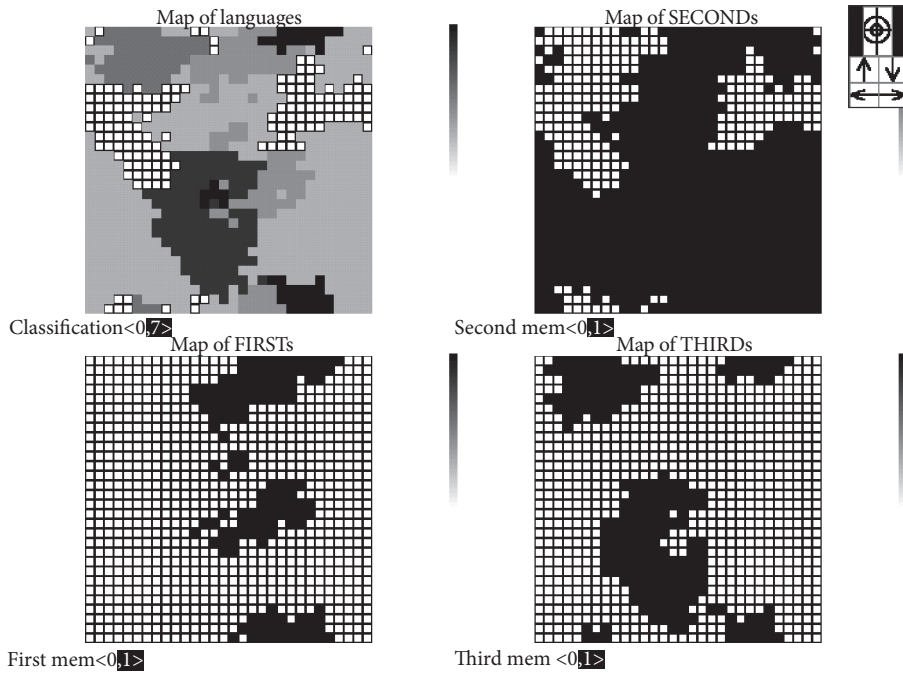

FIGURE 8.7 Feature values after step 69, showing clustering.

172   *Constructions, complexity, and word order variation*

The display in Figure 8.7 shows the state of each agent for each feature at this point in the simulation. As a consequence of the changes from Figure 8.7, some combinations of feature values increase their distribution over the population, while others decrease. The histogram in Figure 8.8 shows the population levels of the eight languages at this point.

Notice that two languages are dead by this point, i.e. there are no agents with the relevant combination of feature values.

Now suppose that we introduce a complexity bias, such as the scope bias or the dependency bias of §8.5. Then the distribution of properties (in the present case, constructions) in the network may change in a different way than illustrated in Figure 8.7. Crucially, bias introduces an imbalance in the strength of feature values. Suppose that there is a bias in favor of Black. Now, if agent A has White, and interacts with one White and one Black, the bias in favor of Black means that Black now outranks White in the competition. So A will change to Black. Figure 8.9 illustrates, with the bias in favor of Black symbolized as a larger circle.

If the distribution of Black and White in the population is even, as in the present simulation, and the bias in favor of Black is strong enough, there is a good chance that in the end, every agent will switch to Black. How long this convergence on Black takes depends on the strength of the bias and the geography of the network of agents.

Turning back to the question of language variation, I assume that both formal and processing complexity can result in bias. In the case of formal complexity, what is in competition in the social network are the more or less general variants of a construction that express the same meaning, i.e. those with and without arbitrary conditions and exceptions. In the case of processing complexity, what is in competition in the social network are constructions that require greater or fewer computational
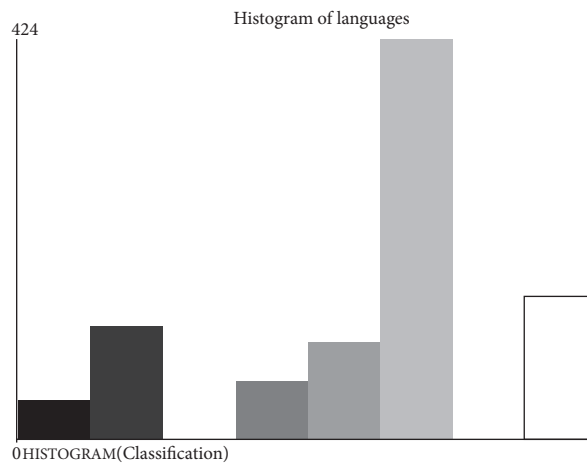

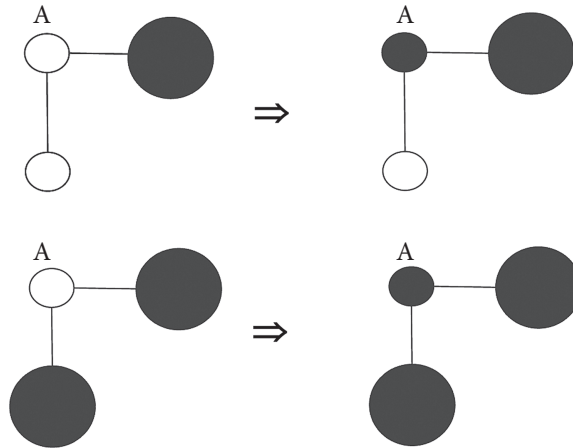FIGURE 8.8 Number of languages of each type.

FIGURE 8.9 Interactions of agents in a network, bias on Black. The state of the network prior to interaction is shown to the left of the arrow, and the state of the network after interaction is shown to the right of the arrow. All interactions are computed and executed simultaneously, so it is possible that an agent may influence one agent but itself be influenced to change by other agents.

resources to construct a given meaning. On the view outlined here, the more general will win out over the less general, other things being equal. And the less complex in processing terms will win out over the more complex, other things being equal.

But other things are not always equal. A cluster of agents may resist external pressure to change, even when the property that is being maintained has a bias against it. The resistance to change may impede the loss of the dispreferred option. I illustrate this point with another computational simulation. I assigned a small negative bias (-1%) to one value (FIRST=Black). The initial state is that typified by Figure 8.10. Note that FIRST=Black occupies roughly the right half of the area, and THIRD=Black occupies the left half of the area. A picture of the state of the simulation after 453 steps is given in Figure 8.11.

At this point each construction has been acquired by a few speakers of the other construction.

After 2200 steps the situation is still stable, in the sense that the area occupied by FIRST=Black (the right half of the area) is still fairly large. However it is smaller than it was before, as shown in Figure 8.12, due to the negative bias. At the same time, the distribution of the THIRD property is essentially unchanged from what it was at the start of the simulation, since there is no bias affecting the distribution of this property.

It is only around steps 3400 to 3800 that the bias ultimately causes the demise of FIRST=Black in the population (Figure 8.13).
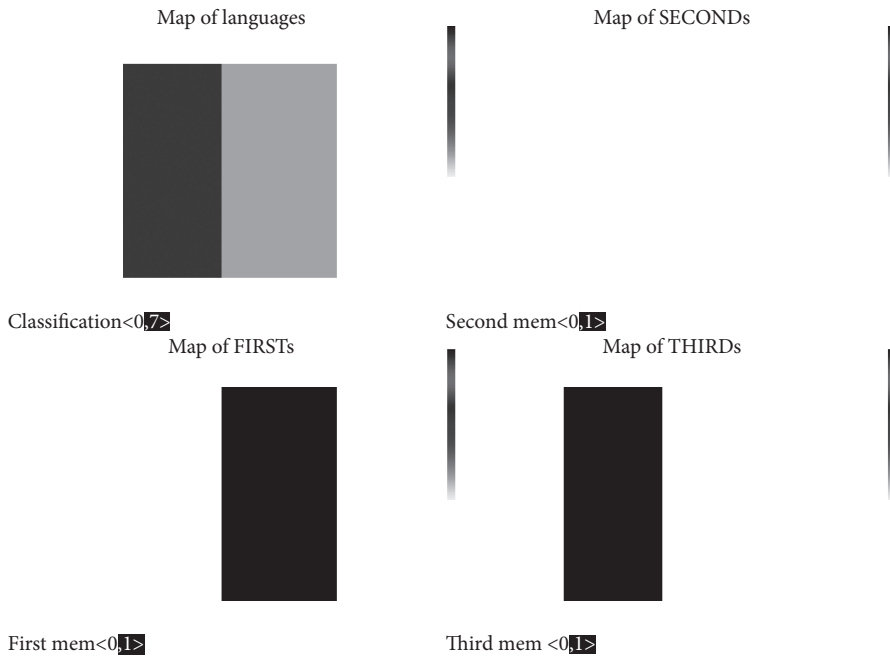
Map of languages   Map of SECONDs

Classification<0,7>   Second mem<0,1>

Map of FIRSTs   Map of THIRDs

First mem<0,1>   Third mem <0,1>

FIGURE 8.10 Two constructions in different regions, initial state.

Map of languages   Map of SECONDs

Classification<0,7>   Second mem<0,1>

Map of FIRSTs   Map of THIRDs

First mem<0,1>   Third mem <0,1>
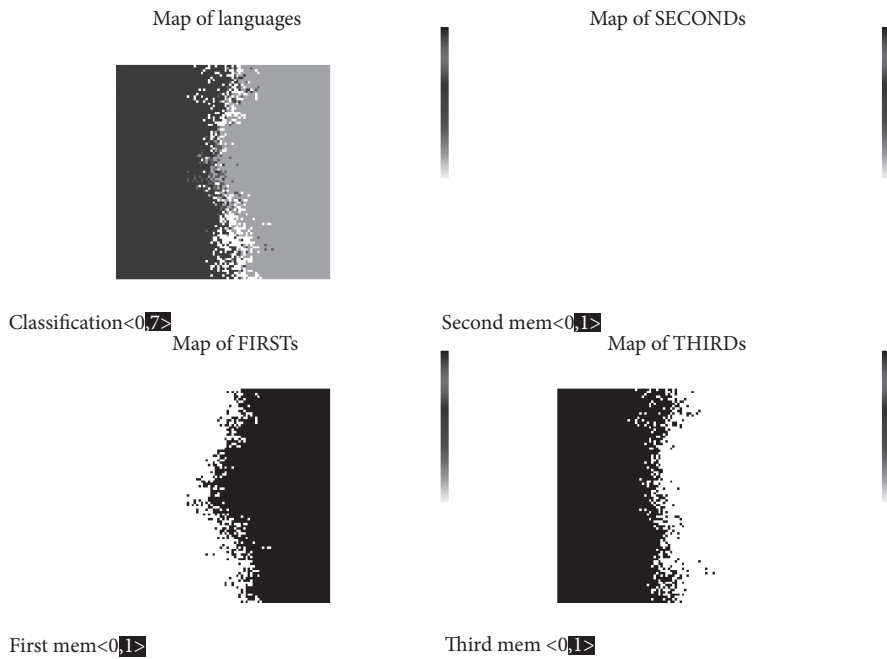
FIGURE 8.11 -1% bias of FIRST=BLACK. Step 453.

Comp. by: PG0844   Stage : Proof   ChapterID: 0002130152   Date:28/4/14   Time:18:29:36
Filepath:d:/womat-filecopy/0002130152.3D175

OUP UNCORRECTED PROOF – FIRST PROOF, 28/4/2014, SPi

Map of languages

Map of SECONDs

Classification<0.7>

Second mem<0.1>

Map of FIRSTs

Map of THIRDs

First mem<0.1>

Third mem <0.1>
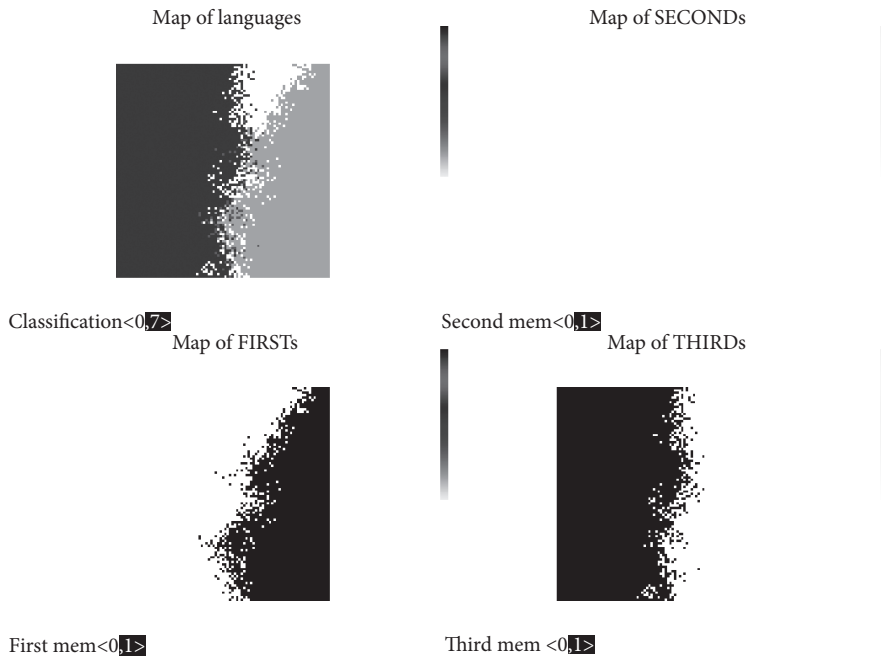
FIGURE 8.12 −1% bias of FIRST=BLACK. Step 2209.

Map of languages

Map of SECONDs

Classification<0.7>

Second mem<0.1>

Map of FIRSTs

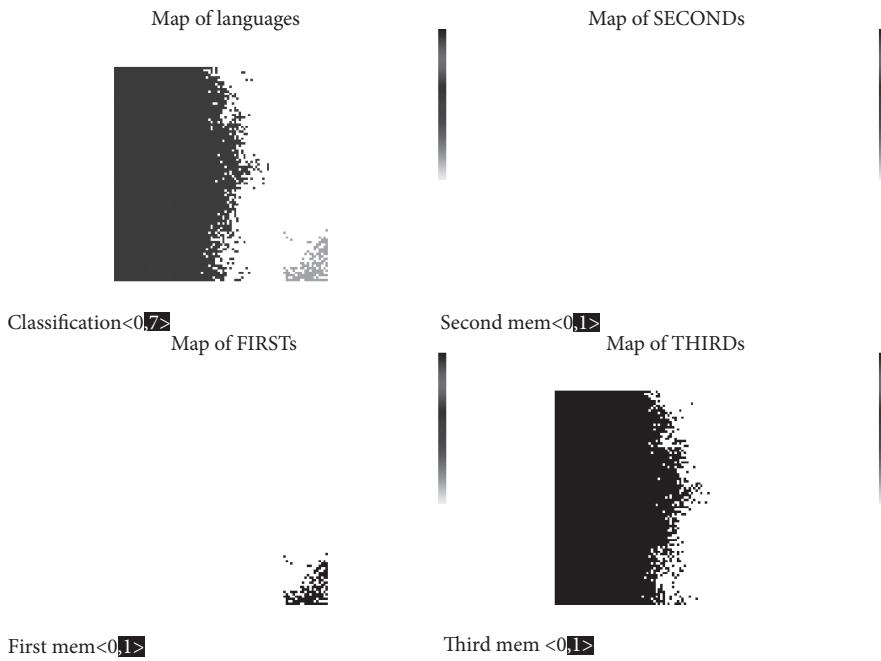Map of THIRDs

First mem<0.1>

Third mem <0.1>

FIGURE 8.13 −1% bias of FIRST=BLACK. Step 375.

[should be 3751]

This simulation illustrates that it is possible for a less preferred option, in this case a particular construction, to remain in the population for a substantial amount of time, and in principle forever, as long as there are conditions that continue to support it. One such condition is lack of contact of agents who have the construction with agents who have competitor constructions; another is a strong and compact cluster of agents that share the same constructions.

Thus a tight knit or relatively isolated cluster of agents may resist the effects of a bias. An agent on the edge of the cluster may change its property value as a consequence of influence of agents outside the cluster. But there is strong support from inside the cluster for that agent to change back to the predominant value in the cluster. Moreover, the agents in the center of the cluster are more resistant to influence from outside, simply because they are more insulated from external influence—they interact with fewer agents who live outside the cluster. So the influence of the bias is less, to the extent that there are well populated clusters that can resist it. The bias adds strength to a particular value, and therefore may induce an agent with another value to switch to it. But what also contributes to the strength of a construction is the number of agents that have it, compared with the number that have the other value.

Finally, what is not reflected in the computational simulation but almost certainly has a significant effect in real social networks is the frequency of use of a dispreferred alternative.[18] This last factor is a plausible explanation for the persistence of irregular verbal morphology, for example, in languages such as English or German.

The computational simulation in this section illustrates a number of points. First, other things being equal, a complexity bias can affect the distribution of properties represented in the social network. Given two constructions of different complexity, the more complex construction is likely to be less represented in the population than the less complex construction. Second, other things are not always equal, and the configuration of the network affects how agents interact with one another. This in turn affects how properties will be propagated in the network. Third, a weak complexity bias may be resisted by clusters of agents, so that a more complex construction may coexist in the network with another less complex construction, perhaps indefinitely.

8.7 Summary and conclusions

In this chapter I have suggested a way of accounting for word order and word order variation in terms of complexity. I have argued that it is useful to consider word order variation in terms of competing constructions. Other things being equal, the less complex construction is preferred to the more complex alternative. When there are

[18] For an extensive examination of the role of frequency in explaining regularity and irregularity in language, see Bybee (2007).

multiple factors contributing to complexity, one construction may predominate in one language variety and a competing construction in another language variety as a consequence of social factors.

I discussed two types of complexity: formal complexity, which has to do with how regular a language is, and processing complexity, which has to do with the computational resources required to relate syntactic configurations and their interpretations. Both types of complexity play a role in language change and variation. However, although it is commonplace to assume that change occurs in order to reduce complexity, I showed that using this notion to understand actual cases of change is far from straightforward. Using the example of English infinitival relatives, I argued that in the case of formal complexity, a change in the grammar that increases regularity in one respect may introduce idiosyncrasy in another respect.

I then turned to Continental West Germanic verb clusters to illustrate how processing complexity may contribute to word order variation. In this case, I argued that it is useful to consider each cluster with a particular order of verbs as a phrasal construction, that is, a form/meaning correspondence. By assumption, all orderings of the daughters of a phrase are available to speakers, although they will actually produce and fully accept those for which there is substantial positive evidence.

Furthermore, I showed that different orderings of the verbs in a cluster differ in complexity with respect to different measures. Those clusters in which the operators, that is, the elements with wide scope, precede their arguments are less complex with respect to the scope bias. Those clusters in which the arguments precede the operators are less complex with respect to the dependency bias, since in these cases, the lexical heads are closer to their dependents.

Given that there are multiple dimensions of complexity, there is no single verb cluster that is least complex. Rather, there are simultaneous pressures to maintain a particular ordering and to change to another order. The availability of all orderings allows for this latter possibility, even where the alternative order is not attested in the language. In a computational simulation, I illustrated how the biases might operate within a social network to create subgroups of speakers who favor one or another alternative constructional variant.

Of course, the situation in the real world is much more complex than is conveyed by the particular cases discussed here. For one thing, there are many factors that enter into word variation, not just the scope bias and the dependency bias. For example, Wasow (2002) shows that word order variation in the English VP is sensitive to grammatical weight, discourse newness, and collocation frequency, among other factors. It is possible that some if not all of these factors are also responses to complexity, on dimensions other than the scope bias and the dependency bias.

Finally, I argued that it is useful to characterize each of the possible verb clusters as an individual construction. In general, it is not possible to say that there is a

Comp. by: PG0844   Stage : Proof   ChapterID: 0002130152   Date:28/4/14   Time:18:29:36
Filepath:d:/womat-filecopy/0002130152.3D178

OUP UNCORRECTED PROOF – FIRST PROOF, 28/4/2014, SPi

178   *Constructions, complexity, and word order variation*

particular ordering of all of the two- and three-verb clusters of a particular variety. Rather, the evidence shows that actually occurring verb clusters must in some cases be defined in terms of specific verbs or verb classes. In the case of Swiss German verb clusters, for example, particular orders (e.g. 2–1) occur in one dialect only when the second verb is AUX. The Swiss German case is particularly interesting because it can be seen that in neighboring dialects, the 2–1 order is extended to clusters in which the second verb is MOD, and in some dialects, to all two-verb clusters. Both the lexical idiosyncrasy of particular clusters and the generalization of particular clusters to more inclusive lexical classes has a natural characterization in terms of constructions.

Comp. by: PG0844   Stage : Proof   ChapterID: 0002130152   Date:28/4/14   Time:18:29:36
Filepath:d:/womat-filecopy/0002130152.3D178

OUP UNCORRECTED PROOF – FIRST PROOF, 28/4/2014, SPi