

Chapter 5

Conclusion

Contents

5.1 Introduction	253
5.2 Summing up	254
5.3 On stochastic grammar	258
5.3.1 Methodological issues	258
5.3.2 Extragrammatical factors determine statistical preferences	263
5.3.3 The multiplicity of genres	269
5.3.4 The reification of style	271
5.3.5 In sum	274
5.4 Future prospects	274
5.4.1 Extending the formal specification of StOT-LFG	274
5.4.2 Extending the case studies	276
5.4.3 Open questions about change	278

5.1 Introduction

In this conclusion, I summarize the core developments of this thesis, and then turn to some foundational issues faced by the approach advocated for here, and end with prospects for future work.

5.2 Summing up

This thesis started from the assumption that syntactic change or cascades of syntactic changes can be interpreted in terms of changes in grammars. Broadly speaking, the thesis was about how syntactic change should be accounted for within a formal model of grammar. To be concrete, it asked a question about the early English changes partially and schematically described in (5.1): (5.1a) is one aspect of the decline of right-headed structures (Chapter 3) and (5.1b) is one aspect of change and variation in the syntax of subjects (Chapter 4).

- (5.1) (a) $OV > VO$ (the decline of right-headed verb phrases)
 (b) $[TOP V NP \dots] > [TOP NP V \dots]$ (the decline of subject-verb inversion in topicalization contexts)

These changes pose difficult problems both at the level of accounting for what happened and at the level of understanding why the changes took place. The aim of this thesis was to ask how changes like those in (5.1) should be accounted for in a formal model of syntax. (5.2) lays out the main issues that were addressed in previous chapters. (5.2d) is a traditional observation in historical linguistics — i.e., change is gradual — and is captured by all accounts of syntactic change. Issues (5.2a-c) are generally overlooked, though.

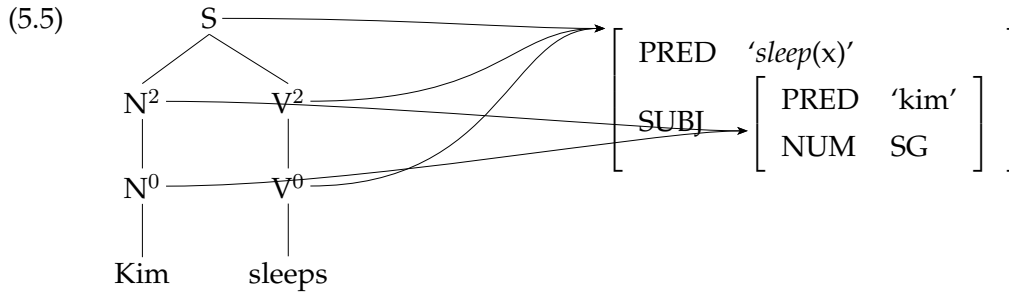
- (5.2) (a) *Structured variation*: Despite extraordinary intratextual variation, certain logically possible word orders are unattested (e.g., the reverse-brace construction) and these unattested word orders reflect typological generalizations (addressed in Chapter 3).
 (b) *Stochastic generalization for change*: Categorical phenomena at one stage of the language show up as preferences at another stage (addressed in Chapter 4).
 (c) *Intratextual variability*: Individual texts display variation at each stage (addressed in Chapters 3 and 4).
 (d) *Time span*: Syntactic change takes place over a span of hundreds of years (addressed in Chapters 3 and 4).

In this thesis, I argued for a novel grammatical model for syntactic change, StOT-LFG,

that provided a new solution to (5.2d), while also accounting for issues (5.2a-c). The core proposals of this thesis are given in (5.3). Individual grammars are taken to consist of constraints (5.3a). Models of these constraints are the types of structures that linguists are familiar with (e.g., finite ordered trees, directed graphs). Different rankings of these constraints determine typological patterns of word order change and variation (5.3b). With respect to syntactic change, the rise in the use of an innovation is reflected by the reranking of the relevant constraints along a continuous scale of real numbers (5.3c). Intratextual variation at each stage of a change is modeled in terms of a nondeterministic component in the grammatical model (5.3d) (Weinreich et al. 1968; Boersma 1998).

- (5.3)
- (a) Grammars consist of sets of axioms — constraints — stated in a formal logic with a model-theoretic interpretation.
 - (b) Patterns of possible and impossible word order change and variation are understood in terms of different rankings of structural markedness constraints, hierarchical constraints, and alignment constraints (Sells 2001c).
 - (c) The steady quantitative rise in the rate of use of an innovation is reflected by gradual constraint reranking along a continuous scale of real numbers.
 - (d) The grammatical model includes a nondeterministic component that captures intratextual variability at different stages of a syntactic change.

In Chapter 2, I presented a description language, L^{lfg} , that was used as medium for stating constraints on LFG analyses. The analyses are triples consisting of a c-structure, an f-structure, and a correspondence relation ϕ , as in (5.4). An example analysis is given in (5.5). L^{lfg} is a precise, extensible description language for expressing violable and inviolable constraints on syntactic structure.

(5.4) $\langle \mathbf{c\text{-}structure}, \phi, \mathbf{f\text{-}structure} \rangle$ 

The description language L^{lfg} has a couple of important properties, listed in (5.6). First, formulae of L^{lfg} are evaluated inside analyses like (5.5) at a particular node or point. It is this property of L^{lfg} that makes it a good formalism for OT. Second, by using a description language of this type, we are able to bridge the formal divide between classical constraints on LFG c-structures and Optimality-Theoretic syntactic constraints (Kuhn 2003b, 94ff.). L^{lfg} provides a unified formal language for stating both violable and inviolable constraints on c-structure.

(5.6) **Properties of the description language L^{lfg} :**

- (a) Local constraint satisfaction (Blackburn et al. 2001, 18)
- (b) Unified formal language for stating violable and inviolable constraints on c-structure, f-structure, and the c-structure/f-structure relation

Having touched on the core formalism of the framework, I now turn to some properties of the StOT-LFG approach to syntactic change, teasing out, in particular, some significant implications for the field of diachronic syntax. Some of these properties were discussed in Chapter 2.

First, within the StOT-LFG approach to syntactic change, smooth changes in the relative frequencies of usage are predicted. This reflects the fundamental observation that (almost) all changes involve layering: if β replaces α , then there must be an intervening stage in which α and β are in variation (Hopper and Traugott 1993, 36; Harris and Campbell 1995,

48-49; Croft 2000, 49-50). According to the analysis in Chapter 3, the decline of right-headed structures in early English is modeled in terms of the gradual promotion of HEAD-L vis-à-vis SPINE-R along the continuous scale of real numbers. Similarly, in Chapter 4, as CANON-X' is gradually promoted with respect to *XP, subject pronouns begin to share the distribution of full NP subjects in Middle English. In contradistinction, categorical approaches to grammar must posit competing grammars (Kroch 1989b) or adaptive rules (Andersen 1973) to account for the instances of gradual syntactic change discussed in this thesis.

Second, StOT-LFG (paired with a learning algorithm like the GLA) yields a dynamical system that matches attested trajectories of change. For example, StOT-LFG accounts for the 'S'-curve for change: change starts out slowly, progresses quickly at midcourse, then slows down in the last stages of the change ('slow-quick-quick-slow'; Aitchison 1981, 97). As discussed in Chapter 2, if the difference in ranking of two constraints which are crucial to the choice between two outputs A and B are changing linearly, then the proportion of output A yields an 'S'-curve.

Third, in the StOT-LFG approach to syntactic change, there is no pressure for grammatical systems that exhibit multiple options to disappear over time. This seems to fit the facts. For example, variation between relative clauses with a relativizer (e.g., *this is the cake that I made*) and bare relatives (e.g., *this is the cake I made*) goes back to Old English (Traugott 1992, 228; Jespersen 1954, 81,132-153). In contrast, Kroch (1994) suggests that grammatical variation is formally analogous to the situation in which a single form has two morphological realizations (e.g., *dived/dove*), which Kroch points out is unstable over time. Hence, the situation of grammatical change is unstable, and must progress toward a single grammatical option. Under the view adopted in this thesis, there is no similar analogous pressure for the stochastic aspects of the ranking to disappear over time. This may not be right for all types of change, though, as evidenced by cross-linguistic directionality in change for certain syntactic phenomena. In Section 3.3.1, I discussed how the StOT-LFG could be extended to account for the typological preference for consistent branching.

Finally, StOT-LFG provides a direct connection between categorical constraints at one stage of a language and statistical preferences at another; i.e., the stochastic generalization for change. In contrast, frequentistic phenomena such as we see during periods of change belong to choices among alternative dialect grammars in the competing grammars approach. In Chapter 4, the stochastic generalization for change was illustrated by change

and variation in the syntax of subject pronouns. I showed that StOT-LFG was able to connect a (near) categorical restriction on subject pronouns in Old English and early Middle English (namely, that they appear outside the verb phrase) to a soft preference on the distribution of subject pronouns in late Middle English.

In the remainder of this chapter, I begin by evaluating a central assumption of this thesis: the idea that probabilistic information should be part of a model of linguistic competence. Following that, I give some prospects for future work.

5.3 On stochastic grammar

A guiding assumption of this thesis is that linguistic competence accommodates and generates variation, and includes a quantitative, noncategorical, and nondeterministic component (Weinreich et al. 1968; Bender 2001). In this section, I examine some of the implications of incorporating probabilistic information into models of linguistic competence (henceforth, *stochastic grammar*).

More generally, what follows is a defense of the *inherent variability* tradition of modeling linguistic variation.¹ As discussed in the introduction, the inherent variability tradition includes the variable rules approach (see references in Paolillo 2002), classification and regression tree analysis (Breiman et al. 1984, Ernestus and Baayen 2003), analogical modeling of language (Skousen 1989, Ernestus and Baayen 2003), generalized linear models (see references in Manning 2003 and Paolillo 2002), various versions of Optimality Theory (e.g., Stochastic Optimality Theory (Boersma 1998), partial ordering (Anttila 1997), floating constraints (Nagy and Reynolds 1997)), extensions of HPSG (Bender 2001), and extensions of the Principle and Parameters framework (Yang 2003).

5.3.1 Methodological issues

I will begin with a methodological argument against stochastic grammar, raised by Newmeyer (2003, 696). This argument says that advocates of stochastic grammar typically draw probabilities from corpora that encompass data from a wide variety of speech communities: “But how could usage facts from a speech community to which one does not belong have any relevance whatsoever to the nature of one’s grammar” (2003, 696). This argument

¹See Abney (1996), Manning and Schütze (1999), Wasow (2002), and Manning (2003) for related defense of stochastic grammar.

bears directly on the choice of corpora used as the empirical base of this thesis (e.g., the Penn-Helsinki Parsed Corpus of Middle English, Second Edition). Each of these corpora encompasses data from a geographically and temporally diverse collection of speech communities.

On the use of corpora

Is it reasonable to use frequency asymmetries in corpora to justify theories of individual linguistic systems? In the case studies discussed in this thesis I took the frequency asymmetries present in individual texts at face value. One potential problem with this approach is that some of the syntactic variants I investigated in this thesis (e.g., the all-final construction in early Middle English, discussed in Chapter 3) were at a low frequency. That being the case, these variants were either relics or productive, but low frequency, syntactic options for individual writers. Without access to early English writers, there is no way of knowing for sure.² The methodological approach taken in this thesis was that the first goal of studies of syntactic change is to provide an accurate characterization of the corpus. For example, in Chapter 4, I created grammars for three early English texts with output distributions of different syntactic options that closely matched the frequency distributions present in those texts. Models of individual texts like those discussed in this thesis form the foundation for a complete understanding of the author's linguistic competence (Kroch and Taylor 2001, 137). I take it that the grammar for *Ælfric's Lives of Saints* in Chapter 4 is a (partial) model of the linguistic system of *Ælfric*, the Abbot of Eynsham.

Accurately characterizing individual texts is the first goal of historical studies. Another goal is to find out in what ways and to what extent do individual texts replicate patterns in the larger corpus of historical texts; e.g., by comparing the frequency distribution of different syntactic options in a historical text with contemporaneous texts found in the same dialect area. By pursuing this goal, we can provide support for the claim that infrequent variants are in fact productive options for individual writers. In Chapter 4, I discussed the late Middle English text, the *Wycliffite Sermons*. In this text, subject pronouns inverted with respect to the finite verb in topicalization contexts about 42% of the time. This distribution reflects the distribution of subject pronouns in other texts in late Middle English: although subjects pronouns could invert in late Middle English, there was a stronger preference

²This is simply a restatement of the methodological truism that historical linguists have no choice but to use corpora as the empirical base of their research.

for subject pronouns to appear before the finite verb in topicalization contexts (Haerberli 2002b).

More generally, advocates of stochastic grammars like those discussed in this thesis typically privilege corpora as the empirical base of linguistic analyses. In practice, large corpora can be drawn upon to develop descriptions of the language use of a wide range of speech communities, as in Statistical NLP (Manning and Schütze 1999, 7). The central question in this type of approach is whether the corpora that are available are representative of communities of speakers. This question is addressed by using statistics to handle finite samples of potentially infinite datasets (Brew and Moens 2000).

The latter approach typically does not separate data from different individuals. In order to justify the modeling of corpus frequencies as the possible output of an individual speaker's linguistic competence, we are hypothesizing that speakers share the same linguistic competence (Boersma 2001). By mixing data from different individuals together in a large data set, evidence relevant to the investigation of the linguistic competence of particular individuals is potentially obscured (Mohanani 2003; Newmeyer 2003). For example, in a situation where some individual mental grammars have pattern X (e.g., verb-object) and other individual mental grammars have pattern Y (e.g., object-verb), combining data from these two types of linguistic systems results in apparently random free variation (Mohanani 2003).

There is no reason, however, to privilege data from the production of an individual speaker or writer, as the methodological argument cited in the introduction to this section (Newmeyer 2003, 696) does. While the tack taken in this thesis was to model the production of individual early English writers, it is often methodologically necessary to look at groups of speakers. In practice, both corpus studies and psycholinguistic experimentation (e.g., careful elicitation of acceptability judgments) absolutely depend on pools of speakers to get statistically significant results; see recent practical discussions of the use of scientific methods in syntax in Cowart (1997) and Schütze (1996).³

In studies of **ongoing** syntactic change and variation, a possible approach is to correlate acceptability judgments with the relative frequency of variants (Anttila 1997, 12 and

³Further, small interconnected groups of speakers (i.e., social networks) contribute to linguistic change and history. Pratt and Denison (2000) discuss the Southey-Coleridge circle in 1790s. They hypothesize that the use of the progressive passive (e.g., *His tooth was being pulled out (by a barber)*; Pratt and Denison 2000, 411) by this social network might have influenced the spread of this construction in English historical syntax.

Mohanan 2003).⁴ For example, consider a situation where a certain variant X (e.g., object-verb) is less frequent than another variant Y (e.g., verb-object). If variant X is also less acceptable for an individual speaker than variant Y, then we might have good evidence for attributing the frequency asymmetry to a property of linguistic competence.

This is not enough, though: “we still have to factor out effects of other language use factors such as comprehension difficulties, direct effects of frequency on acceptability judgments, and so on” (Mohanan 2003). Boersma (2004) discusses mismatches between acceptability judgments and corpus frequencies, noting cases from phonology and syntax where corpus frequencies of a given variant are larger or smaller than acceptability judgments predict.⁵ He argues that these mismatches can be explained by differences between the linguistic task of production and the task of providing acceptability judgments. For production, the speaker chooses a pronunciation that best harmonizes speaker- and listener-based requirements. In contrast, acceptability judgments involve choosing, for a given meaning, an overt form that comes closest to realizing it. Mismatches between acceptability judgments and production occur because of the role speaker-based preferences play in the production process. For example, corpus frequencies of Heavy NP-shift (e.g., *I loaned to Kim the movies that I rented*) are higher than acceptability judgments predict (Boersma 2004, citing John Hawkins, p.c.). The high rate of Heavy NP-shift may be partly rooted in a speaker-based constraint on utterance planning (Wasow 2002).

Summing up, the force of the methodological argument is undermined once we clarify what corpora can be employed for. Further, it is, in fact, often methodologically necessary to look at groups of speakers or writers to get statistically significant results. Lastly, recent work suggests how one can use the frequency asymmetries in corpora to get at the nature of individual linguistic competence, while avoiding the postulation of community grammars.

In the next section I turn to a related methodological argument directly connected to

⁴Acceptability judgments are a type of behavior or performance where “an acceptable sentence is consciously accepted by a speaker as part of his or her language” (Schütze 1996, 20).

⁵Ultimately, we would like to find out whether or not we see differences between the frequencies of a given variant in spoken and written corpora and how these differences match up with acceptability judgments. It could be the case that acceptability judgments converge with spoken, but not written, corpus frequencies for a given syntactic variant. I overlook the crucial distinction between the modality of different corpora here and assume that, in principle, acceptability judgments can diverge from corpus frequencies regardless of the spoken or written nature of the corpus.

the approach taken in this thesis. This argument bears on the relationship between the frequency distributions present in corpora and the frequency distributions output by stochastic grammars.

Overtraining

In Chapters 3 and 4, I tested the StOT-LFG model on two sets of data: (a) frequency data for early English clause structure orderings and (b) frequency data for the distribution of early English subjects. For these two data sets, the Gradual Learning Algorithm closely was used to match the frequency distribution generated by the learned grammar with the frequency distribution for the learning data. This close match was indicated by the low average error rates per candidate.

In Statistical NLP, the first step of testing a model is to achieve a good fit between the training data and the predictions of the grammar (Keller and Asudeh 2002, 230; Manning and Schütze 1999, 206). The second step is to test the model on *unseen* (or *test*) data. This second step helps us to assess how well the model works.

For both the data sets in Chapters 3 and 4, the first step was performed. However, I did not perform tests on unseen data. Keller and Asudeh (2002, 230) present this as a criticism of the research presented in Boersma and Hayes (2001). Boersma and Hayes tested the StOT (+ GLA) model on several data sets, including frequency data for Ilokano reduplication and metathesis, frequency data for Finnish genitive plurals, and acceptability judgment data for English light and dark /l/. Without testing on unseen data, Keller and Asudeh argue, it could be that the grammars generated from the training data are *overtrained*: the model closely matches the learning data but cannot generalize to unseen data (Keller and Asudeh 2002, 231). This is quite serious for at least one of the cases discussed in this thesis: five constraints are used to capture six data points with respect to early English (Chapter 3).

While testing on unseen data is common practice in Statistical Natural Language Processing, it is not entirely clear what data would be tested on for the phenomena in Chapters 3 and 4. The number of data points for each of the data sets are very small. In this case, a possible test for overtraining would be the *leave-one-out-procedure* (Keller and Asudeh 2002, 232). For this procedure, the model is trained on all the data points in the training data, leaving out only one item. The resulting model is then tested on this remaining item. This method is repeated k times (where k is the number of data points). The average model is

then computed. It is also possible to test on tokens of the given training examples, using the same procedure. The advantage of this procedure is that it suits small data sets like those discussed in this thesis (Jäger 2003).

But would testing on unseen data prove anything? It would not say much about StOT-LFG (with the GLA) as a framework for modeling language change and variation. Overtraining is dependent mainly upon the set of constraints and the size of the training data (Jäger 2003; Manning and Schütze 1999, 206-207). For example, a small set of training data may not be “indicative of the complexity of the language at hand” (Manning and Schütze 1999, 206-207). The learning algorithm as such has little relation to either of these two things. However, testing on unseen data would allow us to see whether or not there are too many constraints in the models of variation discussed in this thesis; e.g., the theory of subjects discussed in Chapter 4. I leave this exploration to future work.

In sum, in this section I rebutted a methodological argument against using corpora that represent a wide variety of speech communities to say something about the nature of linguistic competence. Further, I discussed how to address the overtraining problem for stochastic grammars.

5.3.2 Extragrammatical factors determine statistical preferences

The next argument I address critiques the evidence for incorporating probabilistic information into models of linguistic competence (Newmeyer 2003, 696-7). The evidence is weak, goes the argument, because the numbers are epiphenomenal. Consider extraposition of sentential subjects, as in (5.7) (examples from Manning 2003, 307):⁶

- (5.7) a. It is unlikely that the company will be able to meet this year’s revenue forecasts.
- b.#That the company will be able to meet this year’s revenue forecasts is unlikely.

According to this argument, it is wrong to say that, because speakers are more likely to say (5.7a) than (5.7b), this is part of knowledge of language. Rather, speakers tend to avoid sentences like (5.7b) because they are more difficult to process. Apart from parsing

⁶The # mark is used by Manning to indicate that people prefer to put *that*-clauses after the finite verb in clauses like (5.7).

ease, other factors that are taken to influence statistical preferences in language use include real-world knowledge and behavior, and context.

The stochastic generalization

For proponents of the argument just cited, preferences like (5.7) reflect processing, rather than grammar-internal, factors (Newmeyer 2003, 36). For example, the asymmetry illustrated in (5.7) could be explained by pressure to shorten recognition time for phrasal constituents (Hawkins 1994).⁷ Whether the factors that influence grammatical preferences are extragrammatical or not needs to be decided on a case-by-case basis. Many things can influence the choice of one variant over another; e.g., memory limitations. However, it does not follow from the fact that many things influence the choice of a syntactic variant, that knowledge of the frequency of use of certain constructions is not and should not be part of models of linguistic competence.⁸

Further, factors that influence the choice of variants in one language can often be found to have a categorical influence in other languages; see Givón (1979, 26-31), Dik (1997, 34ff.), and Bresnan et al. (2001). As discussed in Chapter 2 and Chapter 4, this observation has been called the *stochastic generalization*. For example, Bresnan et al. (2001) show that just as the person hierarchy affects subject selection *categorically* in Lummi, it also affects the *frequency* of subject selection in active/passive choices in English. In Chapter 4, I discussed an example of the stochastic generalization for change: categorical constraints at one stage of a language can show up as statistical preferences at another.

The thrust of the argument from the stochastic generalization can be further illustrated by considering an instance of the Principle of End Weight in English, discussed in Wasow (2002). The Principle of End Weight says that *phrases are presented in order of increasing weight* (Wasow 2002, 3). The Principle of End Weight arguably holds for the canonical constituent order of English, as illustrated in (5.8) (slightly adapted from Gazdar and Pullum 1981, 120, cited in Wasow 2002, 4). Since heads are single words, they are less complex than phrases.⁹ Prepositional phrases can contain noun phrases as a proper subpart, so

⁷See Wasow (2002, 9-12,25) for discussion of how extraposition is influenced by the weight of constituents and for different conclusions than Newmeyer (2003) about the implications for theories of linguistic competence.

⁸Thanks to Christopher Manning for discussion of this point.

⁹Number of words (vs. phrases) is only one measure of complexity. Others are possible; e.g., number of syllables. Under that metric, *elephantiasis* would be more complex than *eyes of newt*. Thanks to Brian Joseph for discussion of this point.

prepositional phrases can be heavier than noun phrases. Further, clauses and verb phrases are heavier than prepositional phrases.

- (5.8) $H < N'' < P'' < V''$
 (Head < Noun phrase < Prepositional phrase < Verb phrase or clause)

Non-categorical phenomena like Heavy NP-shift are exceptions to the canonical order in (5.8). In Heavy NP-shift (e.g., *I loaned to Kim the movies that I rented*) a non-NP sits between a verb and a following heavy NP. Now recall that the argument raised at the beginning of this section (due to Newmeyer 2003, 696-697) assumes that, on the one hand, in language use people are sensitive to extragrammatical factors such as processing preferences, context, and real world knowledge and behavior. On the other hand, linguistic competence consists of only categorical information. Wasow (2002, 139) argues that this division of labor misses a generalization:

... the Principle of End Weight, which depends on relative measures of length and complexity, manifests itself in paradigmatically grammatical phenomena: the canonical constituent order of English (and other languages), and the obligatoriness of one form in certain syntactic alternations when the direct object is a pronoun [e.g., the verb-particle construction — BZC]. If the canonical ordering and the obligatory cases are part of competence grammar, but the quantitative preferences are treated as performance, then a larger generalization is lost.

A similar argument is made by Givón (1979, 26-31) (cited in Bresnan et al. 2001), Bresnan et al. (2001), and Bresnan and Nikitina (2003).

It is possible to argue, as Newmeyer (1998, 39-44) does, that there is no missed generalization: a single external factor exerts its influence on language users, which has categorical effects in some languages – where the results of the force are part of linguistic competence, but not in others – where speakers have grammatical options.

Newmeyer's viewpoint can be characterized in the following way; see Newmeyer (1998, 39-44) for related discussion. Take two languages, L_1 and L_2 . In L_1 (e.g., English), both indefinite and definites are possible subjects. However, indefinite subjects are rare. In L_2 (e.g., Bemba (Givón 1979) and Malagasy (Keenan 1976)), indefinite subjects are prohibited. The unifying external factor that connects the distributions found in L_1 and L_2

is that there is a general communicative preference to express nominal expressions with contextually salient referents (e.g., definite NPs) in perceptually salient positions (subject or clause-initial position).

As a counterpoint, consider the OT account of the interaction of definiteness and grammatical role in Aissen (2003). The constraint subhierarchy in (5.9) forms part of Aissen's account. The two highest ranking constraints, *Su/NSpec and *Su/Spec, prohibit non-specific indefinite and specific indefinite subjects, respectively.

(5.9) *Su/NSpec \gg *Su/Spec \gg *Su/Def \gg *Su/PN \gg *Su/Pro

The subhierarchy in (5.9) is grounded in cognitive/communicative tendencies. Where Aissen's account differs from Newmeyer's is that the subhierarchy in (5.9) is claimed to be part of the synchronic mental linguistic systems of speakers of both L₁ and L₂. In L₁, where the constraint subhierarchy is low-ranked, the effects of the constraints can be detected in only a restricted set of contexts (Bresnan and Aissen 2002a, 88). In L₂, the constraint subhierarchy is highly ranked and the constraints have a categorical effect (e.g., in Bemba and Malagasy).

Both Newmeyer's account and Aissen's OT account postulate that there is a unifying explanation for the distribution of indefinite and definite NPs. In Newmeyer's account the explanation is to be found in external factors (communicative preferences), whereas in the OT account, the explanation is to be found in an internal factor (a constraint subhierarchy), albeit one grounded in communicative preferences.

I now turn briefly to some phenomena that I think favor the OT account over Newmeyer's account. There are a growing number of analyses of phenomena that appear to demonstrate that low-ranked and hence relatively inactive constraints are actually present in the synchronic state of individual mental linguistic systems. For an account like Newmeyer's, this type of phenomenon would have to be explained by extragrammatical factors; e.g., processing constraints. What is needed then is an explicit proposal about what these extragrammatical factors are and how we label them as such. I will discuss two phenomena that pose a challenge for Newmeyer's account. For both of these phenomena, I make reference to language users reranking constraints. The relevant behavior for both of these cases could have occurred for various reasons; e.g., accommodation through the elimination of marked structures which are inaccessible to interlocuters (Bresnan 2000b). This behavior is modeled in terms of the elevation of constraints. There is no sense in which the reranking

of constraints *causes* the relevant behavior; cf. McMahon (2000).

The first case involves the pronunciation of non-native forms. Smolensky et al. (2003, 19ff.) demonstrated that when faced with non-native inputs, speakers can elevate faithfulness constraints from their base positions to higher positions. Smolensky *et al.* asked adult English speakers to produce forms with initial clusters like [kt] that are illegal in English. They found that for some forms, speakers were able to produce faithful, English-illegal outputs, with no nativization. On this basis, they proposed that when producing these English-illegal forms, speakers prioritize low-ranked faithfulness constraints above the relevant markedness constraints. This example suggests, contra Newmeyer, that constraints that are ranked too low to have a categorical effect in a language are actually present in synchronic mental linguistic systems.

The second case involves pidgin genesis. In situations where the contact languages are typologically distant (i.e., share few marked forms in common), creators of pidgins rerank low-ranked individual markedness constraints above conflicting faithfulness constraints (Bresnan 2000b). Reranking the constraints in this way has the effect of removing marked forms (e.g., bound pronouns). Bresnan (2000b) discusses markedness of pronoun forms. She assumes two types of constraints. First, a set of markedness constraints against complex or difficult structures; e.g., constraints against pairing bound and zero forms with pronoun semantics. Second, a set of faithfulness constraints requiring that features of the input content be preserved in output forms; e.g., reduced forms are specialized for topical content. In pidgin genesis, simplification is modeled as a process in which speakers eliminate marked features of their language by reranking low-ranked markedness constraints above individual faithfulness constraints that conflict with them. Constraints targeted for reranking are those which mark types of forms that are not understood or not easily learned by interlocutors because they are not in the inventory of the interlocutors' language.¹⁰ For example, if a speaker of a language in which bound forms realize topical, pronominal content is trying to communicate with a hearer that does not have bound forms in his/her language, then the speaker will elevate the markedness constraint against bound forms paired with pronominal content.

In sum, the argument discussed in this section assumes that the factors that influence the choice of syntactic variants are extragrammatical. This runs afoul of the stochastic generalization — factors that affect the choice of variants have a categorical effect in other

¹⁰Hence, Bresnan model of simplification is hearer-based.

languages. An example of the stochastic generalization was discussed in Chapter 4. Further, several recent analyses suggest that low-ranked constraints can be shown to be synchronically active in certain contexts. This type of phenomenon poses a challenge to any approach that attributes the choice of variants solely to extragrammatical factors.

The conflation of competence and performance

The argument that extragrammatical factors influence statistical preferences relates directly to another argument. A guiding assumption of the case studies in Chapters 3 and 4 was that, in cases of optionality, stochastic grammars not only deliver the options, but also predict their frequency of occurrence. According to the argument, this assumption leads to the abandonment of the traditional competence/performance hypothesis because, in order to predict actual frequencies, all the relevant performance factors would have to be put in the grammar; i.e., “all constraints on perception, articulation, memory, fatigue, style, and politeness interact with grammatical constraints” (Keller and Asudeh 2002, 240). For Keller and Asudeh (2002, 240) this leads to absurd conclusions with respect to traditional claims of OT like factorial typology because “surely speakers with distinct native languages have cognitive abilities in common and these cannot be reranked to yield their different languages”. For Newmeyer (2002, 98), this assumption is a short step away from Late Generative Semantics (Lakoff 1974), “which argued that *all* factors relevant to morpheme occurrence in discourse are a matter for grammatical analysis”.

First, as pointed out above, it does not follow from the fact that lots of things (including things that should not be part of a system of constraints) influence the frequency of certain constructions that knowledge of the frequency of use of various constructions is not and should not be part of competence.

Further, it is not true that a commitment to stochastic grammar forces one to abandon the performance/competence distinction. StOT-LFG preserves the classical competence/performance distinction in the following way (Bresnan and Aissen 2002b). Each competence grammar (knowledge of language represented as a set of constraints and their (mean) ranking values) is embedded in a *usage grammar*. I define a usage grammar as consisting of a competence grammar plus some mechanism that allows extralinguistic factors (e.g., politeness, exhaustion) to affect the ranking values of constraints. For some of these factors (e.g., style), a systematic mechanism may be needed to link these factors to particular constraints (see below). For other non-systematic factors (e.g., exhaustion), we need a

model of our ignorance of the entire context and the non-linguistic factors that condition the probability of an output. The *noise* variable of stochastic evaluation may be taken as a model for non-systematic performance effects. However, to conclude that the intraspeaker variability modeled by stochastic grammars is random and uncaused is “the fallacy of reified ignorance” (Bresnan and Deo 2001, 37): “stochastic models represent gaps in our knowledge of the world, not gaps in the causal structure of the world”. Extra-grammatical factors are represented in the usage grammar by variables that perturb the rankings of constraints at evaluation time, **not** as individual constraints.

5.3.3 The multiplicity of genres

The next argument, due to Newmeyer (2003, 697-698), starts from the observation that the probability of use of a syntactic variant is a function of a particular genre or style. However, there are a large number of genres (Biber 1995). According to the argument, stochastic models of linguistic competence are forced to claim that the grammars of speakers who command certain genres differ from others who do not. This is wrong. Instead, we should distinguish knowledge and (appropriate) use, which stochastic grammars fail to do.

The particular example that Newmeyer (2003, 697-698) uses is the “bizarre (but genre-normal) syntax and lexicon” from flight crews. Example (5.10) gives some of Newmeyer’s examples.

- (5.10) (a) We are ready to depart the gate.
 (b) Takeoff will be shortly.
 (c) We hope that you will enjoy your stay in the Seattle area or wherever your destination may take you.

If the examples in (5.10) are part of a small set of memorized formulae, then they are not relevant to stochastic grammar. However, if the examples can be demonstrated to be a reflection of structured variation, then the issue becomes how to model stylistic and genre variation in stochastic grammar.

This argument assumes that stochastic grammars (e.g., via stochastic evaluation, probability distributions over argument frames, etc.) completely derive output distributions. This is problematic: the information encoded in these grammars might be epiphenomenal, because extragrammatical factors, like genre and style, actually determine the probability

of use of a particular syntactic variant. It is this last point that reveals another underlying assumption of this argument: knowledge of language does not include a mechanism which links extralinguistic factors to linguistic variables.

As a starting hypothesis, let's assume that there is no systematic relationship between grammatical constraints and extragrammatical factors like genre and style. Exactly this is reflected in the basic architecture of StOT-LFG, as presented in Chapters 2, 3, and 4. In StOT-LFG, linguistic competence is modeled as the ranking values of individual constraints. Specific occurrences of linguistic variants are modeled as random and unpredictable. These performance effects are modeled by the normal distribution used in the stochastic evaluation of the constraints.

The evaluation point ('selectionPoint') of a constraint C_i at evaluation time is given by the equation in (5.11) (adapted from Boersma 2000, 483):

$$(5.11) \text{ selectionPoint}_i = \text{rankingValue}_i + \text{noise.}$$

There is good reason to believe that this picture is overly simplistic. Weinreich et al. (1968, 169) define a strict condition under which a linguistic variable can be defined: "Quantitative evidence for covariation between the variable in question and some other linguistic or extralinguistic element provides a necessary condition for admitting such a structural unit". One of the central observations of the variationist tradition is that knowledge of language includes sensitivity to linguistic variation that correlates with different styles and genres, *appropriately defined*.¹¹

Variable rules (Paolillo 2002) are one type of mechanism which links extralinguistic factors to linguistic variables. Covariation between linguistic variables and extralinguistic elements, like style and genre, can also be modeled in StOT-LFG by allowing these elements to systematically boost or depress the ranking values of selected constraints. Boersma and Hayes (2001, 83-84) discuss how this might be implemented.¹²

We assume that utterances occur in contexts that can be characterized along a casual-to-formal continuum. We quantify this continuum with a variable *Style*, such that *Style* equals 0 in maximally casual speech and 1 in maximally formal

¹¹Recent work has problematized the classical notion (Labov 1966) of reified style categories; see the papers in Eckert and Rickford (2001) and below.

¹²See Lee (to appear) for discussion of how Boersma and Hayes' (2001) Stochastic Optimality Theory model of stylistic variation can be applied to intraspeaker variation in the ellipsis of case markers in Korean.

speech. The selection point for a given constraint C_i is determined by [the following equation:]

$$\text{selectionPoint}_i = \text{rankingValue}_i + \text{styleSensitivity} \times \text{Style} + \text{noise}$$

This is the same equation as before [(5.11) – BZC], augmented by the term $\text{styleSensitivity}_i \times \text{Style}$, in which $\text{styleSensitivity}_i$ is a constraint-specific value. Constraints with positive values for styleSensitivity take on higher ranking values in formal speech; constraints with negative values for styleSensitivity take on higher ranking values in casual speech, and constraints with zero values of styleSensitivity are style insensitive.

Note that the styleSensitivity factor is not itself a constraint. Rather, it is a variable that systematically boosts or depresses the ranking of specific constraints or groups of constraints. In this way, it makes good on the following “general statement” by Weinreich et al. (1968, 187-188):

Linguistic structure includes the orderly differentiation of speakers and styles through rules which govern variation in the speech community; *native command of the language includes the control of such heterogeneous structures*. [my emphasis – BZC]

Given the model that Boersma and Hayes lay out in the quote above, speakers/writers who command certain styles and genres do not have a different constraint set than those who do not, contra the argument discussed in this section. Rather, systematic covariation between linguistic variables and extralinguistic factors is modeled in terms of weights that alter the ranking of constraints.

In sum, StOT-LFG without a styleSensitivity factor treats the output distribution of linguistic variables as a performance effect. However, one of the major insights of the variationist tradition is that linguistic variables covary systematically with extralinguistic factors like style and genre. The addition of the styleSensitivity factor to the basic architecture of StOT-LFG makes good on this insight.

5.3.4 The reification of style

The final argument I discuss takes issue with the stochastic grammar approach to stylistic variation: stochastic grammar starts from a purely grammatical theory and adds on a

mechanism (e.g., stochastic evaluation) for capturing, as Casillas Martínez (2003) puts it, “the observational result of variation” (frequency distributions either at the level of the community or the individual). For example, in StOT-LFG the choice of variants is purely random, a consequence of a stochastic evaluation. However, according to the argument, stylistic variation is not random: rather, stylistic variation is a social process, where speakers intentionally “choose” variants (resources) to construct styles or personae (Bender 2001; Casillas Martínez 2003).

First of all, it is important to first clarify the role of *noise* in the StOT-LFG framework. As discussed above, the *noise* variable represents unknown factors that are independent of the theory embodied by the constraint set; i.e., noise is a model of our ignorance of the entire context and the non-linguistic factors that condition the probability of a certain output. Further, it is not reasonable to assume that every use of a variable constitutes a socially meaningful choice: it is an open question what the relationship is between more and less intentional uses of variables (Eckert 1996, 58).

Second, the preceding section showed how to extend the StOT-LFG model to account for the interaction of extralinguistic factors and linguistic variables. There is some reason to think, though, that the model of stylistic variation presented in response to the argument discussed in the previous section is overly simplistic. Boersma and Hayes (2001) adopt the classical variationist approach to style (Labov 1966), wherein intraspeaker variation is treated as unidimensional, a function of attention paid to speech, ranging from casual to highly monitored (formal) speech. This traditional approach assumes the stability of style and social categories and limits style to a single dimension (degree of formality), thus failing to explain how other factors contribute to intraspeaker variability (Podesva et al. 2001, 178).

In contrast, recent work on variation¹³ views style as the linguistic means through which identity is constructed in context, possibly understood as a collage of co-occurring linguistic features that together constitute social meaning. Can stochastic grammar be reconciled with this view of style? Recent work on style suggests that a full-fledged stochastic grammar model of stylistic variation must have, at least, the properties in (5.12):

¹³What Eckert 2002 calls the “third wave of variation studies”; see, for example, Podesva et al. (2001) and the papers in Eckert and Rickford (2001).

- (5.12) (a) **Multidimensional:** Constraints should be able to index multiple categories or pragmatic meanings whereas the Boersma/Hayes model treats style as unidimensional. For example, final released stops are associated with intelligent nerdy style, preciseness, prissiness, etc.; see Eckert (2002) for discussion.
- (b) **Indexical:** Some constraints will directly index certain social categories, while others index social categories indirectly through pragmatic meanings in local context.
- (c) **Clustering:** Sets of constraints should be able to cluster together to form styles. For example, high pitch, wide pitch ranges and prolonged /l/s together index a gay style (Podesva et al. 2001).

What are the consequences of this view of style for intratextual variation in historical texts, like those discussed in this thesis? In these cases, the appropriate axis may be genre, not style. That is, it is probably incorrect to impute current notions of identity and style to early English authors.¹⁴ I assume that genre can be loosely defined as a classification of texts based on varieties easily discernible by native speakers, corresponding to differences like subject matter, purpose, mode, and discourse situation (Biber and Finegan 1989, 488; Moesnner 2001).¹⁵ According to this definition, genre is distinct from text type (Biber 1994, 52), where text type is defined on a purely linguistic basis.¹⁶ In this thesis, I classified texts according to genre, as defined by subject matter; e.g., religious sermons. For recent work on the relationship between linguistic innovation and genres, see Nevalainen and Raumolin-Brunberg (2003).

In conclusion, once the role of *noise* is clarified the argument from the reification of

¹⁴This is not to say that identity was unimportant. This is evident in John de Trevisa's famous quote discussing the reception of Northern speech by Southerners: "Al the longage of the Northumres and speicialliche at York is so sharp slittyng and frontyng and vnshape (= harsh, piercing, and grating), that we southern men may that longage vnnethe (= hardly) vnderstonde" (c1385). Trevisa is explicitly contrasting along linguistic lines his Southern identity with the Northern. In fact, it may be more precise to say that identity and genre are intertwined: we develop our identities through genres like teaching or giving sermons.

¹⁵This definition of genre is good as a starting point, although it needs to be articulated more fully. See Bauman (2001) and Macaulay (2001) for recent discussions of genre.

¹⁶In their work, Biber and Finegan have explored a multi-dimensional model of text type variation, where each dimension is characterized by a set of frequently co-occurring linguistic features. After text types have been identified on linguistic grounds, they can be interpreted in terms of author purpose, discourse context, etc..

style loses its force. I have also made some tentative suggestions on how to extend the StOT-LFG model to be consistent with recent work on genre and style.

5.3.5 In sum

I have examined a core assumption of this thesis: knowledge of language accommodates and generates variation, and linguistic competence includes a quantitative, noncategorical, and nondeterministic component. The overarching conclusion is that there are compelling reasons (e.g., the stochastic generalization) to explore models of stochastic grammar. In the next section, I discuss future research building upon the work begun in this thesis.

5.4 Future prospects

This thesis presented a novel formal framework, StOT-LFG, for modeling syntactic change, as well as two case studies on early English clause structure change. There is much more work to be done. In this section I discuss some prospects for future work. First, the specification of the formal framework in Chapter 2 is incomplete in various respects. I discuss two possible extensions. Second, the case studies presented in Chapter 3 and 4 contain various gaps. Third, there are many open questions about change not fully addressed here; e.g., directionality.

5.4.1 Extending the formal specification of StOT-LFG

The two issues that I wish to discuss with respect to the formal framework both bear on interface questions. First, how do we connect the syntactic structures discussed in Chapter 2 to formal semantic interpretations? Second, what is the role of morphology?

The syntax-semantics interface In Section 2.4.3, I defined the input to StOT-LFG analyses as an *f*-structure representing semantic and morphosyntactic content independent of its formal realization as an LFG analysis of the form $\langle c\text{-structure}, \phi, f\text{-structure} \rangle$. This is an oversimplification. Ultimately, a complete formal specification of StOT-LFG must include an explicit formal semantics for its representations of meaning (Kuhn 2003b; Andrews 2004). There are at least two ways to include formal semantics in StOT-LFG. The

first would be to include another level of representation in the input; e.g., a s(ematic)-structure (of the sort described in Dalrymple 2001, 217-229) projected from the f-structure (Kuhn 2003b, 63). Equivalently, a feature SEMANTICS could be included in the input f-structures (Kuhn 2003b, 63). A second, related way would draw upon work on the glue approach to semantic composition (Dalrymple 2001, 229-253; Asudeh 2004). In this vein, Andrews (2004) proposes that the input to OT-LFG analyses consists of two components: an f-description¹⁷ that delimits the range of possible f-structures for the candidate set, and a collection of meaning constructors¹⁸ that determine a meaning. An important next step for the work begun in Chapter 2 would be to combine the model-theoretic approach to syntax described there with a glue logic approach to semantic interpretation.

The syntax-morphology interface In Chapter 3 it was shown that in early English all-final constructions rooted in S (e.g., *you God's commandment keep will*) declined in favor of all-medial constructions rooted in I² (e.g., *you will keep God's commandment*). This change correlates with the loss of case morphology (Kiparsky 1997). In Old English and early Middle English, case morphology, not syntactic position, determines grammatical roles like subject and object. This goes hand in hand with a number of early English phenomena; e.g., subjects were not obligatory in Spec,I² and objects could scramble with each other in the double object construction. The loss of morphological case in Middle English correlates with the loss of these phenomena; e.g., the rise of Spec,I² in late Middle English as the obligatory position for subjects.

This is one example of a broader class of changes in which the realization of morphosyntactic content shifts from morphology to syntax.¹⁹ Condoravdi and Kiparsky (2002) discuss the rise of a composite functional projection in Greek. Vincent (1997) shows that DP emerged historically in Romance. This class of changes suggests that the approach taken in this thesis must be extended to account for the observation that morphology provides an alternative to syntactic structure for morphosyntactic content like grammatical roles and

¹⁷An f-description consists of one or more equations of the form in (i). The equation in (i) says that the f-structure *g* has an attribute TENSE with the value PAST.

(i) (*g* TENSE) = PAST

¹⁸Meaning constructors are pairs consisting of a left-hand side (representing a meaning) and a right-hand side (representing a logical formula over semantic structures that correspond to that meaning) (Dalrymple 2001, 231).

¹⁹See Condoravdi and Kiparsky (2002), Kiparsky (1995), Kiparsky (1997), Vincent (1997), and Deo (2001).

tense. It may be possible to understand competition between analytic and synthetic realizations of *f*-structure input in historical morphosyntax in terms of change in the strength of structural markedness and faithfulness constraints over time; see Bresnan (2000b) and Bresnan (2001a).

5.4.2 Extending the case studies

The goal of this thesis was to show that the appropriate grammatical model for syntactic change is one that incorporates a quantitative, noncategorical, and nondeterministic component. In Chapters 3 and 4, I presented two case studies on early English syntax, demonstrating that gradual syntactic change and variation exhibited at the level of individual speakers are an integral part of a grammatical model for syntactic change. The next step for this work is to explore explanations for why the rate of use of constructions like object-verb order and subject-verb inversion declined in Middle English, and show how the factors that caused these changes fit into an explicit model of syntactic change of the sort described in this thesis. A more complete grammatical model for syntactic change will emerge once the different possible internal and external factors that caused these changes are identified. In this section, I look back at the case studies in Chapters 3 and 4 and describe how this work might be accomplished.

The decline of right-headed clauses in early English In Chapter 3, I discussed diagnostic criteria for identifying object-verb and verb-object orders in early English texts. Hitherto, these diagnostics have only been systematically applied to Old English and (parts of) early Middle English. In order to understand why object-verb order was lost in Middle English, it will first be necessary to apply these diagnostics systematically to (southerly) Middle English texts (e.g., from the Penn-Helsinki Parsed Corpus of Middle English, Second Edition) in order to get a precise quantitative picture of change in the frequency distribution of object-verb and verb-object order at different stages of Middle English. The second step would involve investigating the factors that influence the choice of object-verb vs. verb-object order in these texts and that may have led to the decline of object-verb order in Middle English. Some of the factors that have been identified in the literature and that would guide this future research include genre (i.e., more oral vs. less oral texts (Warner 2000); poetry vs. prose (Foster and van der Wurff 1995)), information structure (i.e., given vs. new information (Foster and van der Wurff 1997)), ease of processing (e.g., the factors

that have been identified in recent psycholinguistic work by Hawkins 1994 and Wasow 2002; see 3.3.1), and second language acquisition (Weerman 1993; Trips 2002). The results of this research could be used to refine the different hypotheses of what caused the decline of object-verb order in Middle English, and to extend the grammatical model of syntactic variation and change described in this thesis; e.g., to capture how external factors like genre and processing interact with the grammatical model in language acquisition and use. One suggestion for how to extend the model to treat style and genre was discussed in Section 5.3.4.

The syntax of subjects in early English In Chapter 4, I discussed variation and change with respect to different types of subjects in early English. In that chapter I showed that full noun phrase subjects variably inverted with the finite verb in topicalization contexts. Examples of an inverted and uninverted full noun phrase subject are given in (5.13) and (5.14).

(5.13) **Inverted full noun phrase subject (topicalization)**

[*Twa mila*] *hæfde* **martinus** *fram his mynstre to Turonian byrig*
 Two miles had Martinus from his monastery to Turonian stronghold
þær se bisceopstol wæs,
 where the cathedral was
 “Martinus had two miles from his monastery to the Turonian stronghold where
 the cathedral was”
 (c. 950-1050, *Ælfric’s Lives of Saints*, [SOURCE: YCOE
 (coelive,+ALS_[Martin]:1198.6764)])

(5.14) **Uninverted full noun phrase subject (topicalization)**

and [*ðas feower godspelleras*] **God** *geswutelode* *gefyrn,*
 and these four gospels God made known long ago
 “and these four gospels God made known long ago,”
 (c. 950-1050, *Ælfric’s Lives of Saints*, [SOURCE: YCOE
 (coelive,+ALS_[Mark]:174.3311)])

One open question is how information structure affects the syntactic positioning of full

noun phrase subjects. Birner (1994) shows that inversion in modern English depends upon the *discourse-familiarity* of the preposed or postposed constituent. Discourse-familiarity is determined by, among other things, prior evocation in the discourse, inferability from prior discourse, and recency of mention. Birner demonstrates that information represented by preposed constituents is typically discourse-old, whereas postposed constituents are typically discourse-new. An important next step in the investigation of the syntactic distribution of full noun phrase subjects in early English would be to test Birner's hypothesis with respect to inverted and uninverted subjects. That is, are inverted full-noun phrase subjects typically discourse-new and are uninverted full-noun phrase subjects typically discourse-old?

For subject pronouns, the next step is to pursue the question raised at the end of Chapter 4: why do subject pronouns gradually begin to share the syntax of full noun phrase subjects in Middle English? At the end of Chapter 4, I proposed that the answer to this question lies in changes in the inventory of pronominal types in early English. Namely, I hypothesized that subject pronouns acquired the syntax of full noun phrase subjects as a consequence of the gradual loss of the *se*-series of referentially independent demonstrative pronouns. In order to test this hypothesis, it will be necessary to show a correlation between the loss of the *se*-series and change in the syntax of subjects pronouns (e.g., change in the ability to invert with respect to the finite verb in topicalization contexts) in Middle English texts.

5.4.3 Open questions about change

The model of syntactic change discussed in this thesis is incomplete in many respects. Along the way, I have raised many more questions about variation and change than I have attempted to answer. In this final section, I briefly discuss two ways in which the model should be extended.

Directionality

In Section 3.3.1, I discussed the issue of directionality in word order change and suggested an extension of StOT-LFG incorporating filtered learning (Kirby 1999). In this section, I briefly return to the issue of directionality and chart a path for future work on this issue within the context of probabilistic models of change and variation.

Claims about directionality in linguistic change are claims about how particular grammatical constructions may change. Schematically, directionality claims say that for two grammatical constructions A and B, A may become B but not vice versa. Directionality is a robust phenomenon that appears in all domains of linguistics. For example, in phonological change, voiced obstruents may become voiceless in final position, but the reverse never happens; e.g., Old High German *rad* > Middle High German *rat* ‘wheel’ (Haspelmath 1999, 587). In syntactic change, lexical categories may turn into functional categories, but not vice versa; e.g., N > P (Old French *chies* ‘house’ > Modern French *chez* ‘with’, ‘at’).²⁰ Even researchers who question whether there are in fact directionality generalizations over observed linguistic changes concede that directionality is “challenging as a *phenomenon*” (Lightfoot 2003, 106) and that “unidirectionality is almost true” (Newmeyer 1998, 277).

I take it that there are robust, cross-linguistic generalizations which hold of linguistic change (Hopper and Traugott 2003; Heine and Kuteva 2002). However, there are no language specific causal principles of language change (Haspelmath 1999; Lightfoot 1999, Lightfoot 2003). Rather, motivations for directional change are external; e.g., parsing and production preferences (Hawkins 1994; Kirby 1999; Wasow 2002). The methodology of seeking non-linguistic, psychological motivations for change goes back to the Neogrammarians (Haspelmath 1999).

The locus of explanation for directionality is controversial. An assumption of most accounts of directionality is that the goal of speakers and listeners is not to improve their language (Boersma 1998; Keller 1994; Kroch 2001; Lightfoot 1999), but rather to complete their joint projects as quickly, clearly, and easily as possible (Clark 1996). One class of accounts appeals to speaker and hearer actions to explain language change and directionality (Keller 1994; Andersen 2001; Traugott and Dasher 2002). I will call this the usage-based account of directionality. On this account, individual actions have cumulative consequences that ultimately lead to directionality in language change. A non-linguistic example of individual actions having an impact over time is that of a path that is cut across grass as a consequence of “people making short-cuts to save time, not with the intention of making a path” (Traugott and Dasher 2002, 36ft, citing Keller 1994). The usage-based account proposes a similar account of directionality in language change. In language use, the cumulative consequences of individual actions may or may not be goal-oriented.

A second class of accounts appeals to language acquisition to explain language change

²⁰See, for example, Hopper and Traugott (2003) and Roberts and Roussou (2003), among many others.

and directionality. I will call this the acquisition-based account of directionality. I have discussed two acquisition-based accounts in this thesis. In Kiparsky's (1996) endogenous optimization approach, directionality is accounted for in terms of an interaction between an underlying preference for uniform directionality of head-complement relations (the "efficient" cause) and the attrition of evidence for OV structures (the "enabling" cause). In Section 3.3.1, I explored the filtered learning account (Kirby 1999; Jäger and Rosenbach 2003). In this approach, a processing preference acts as a filter on the language data that the learner hears. As a consequence, the input to the learner is not raw language data, but a subset. Over time, this filtering has an impact on the distribution of syntactic variants.

There were several problems with the filtered learning account, discussed at the end of Section 3.3.1. First, the model accounts for directionality in terms of only one aspect of language use: rapid parsing by listeners. Second, the model assumes that the learner is endowed with substantial computational power. A theory of the computational capacities of language acquirers that can be incorporated into a model of acquisition does not exist (Yang 2003, 28). Until it does, it is safest to assume that the learner is not endowed with much computational power.

A first step towards a more realistic account in the context of StOT-LFG is to develop a model that incorporates other factors, apart from processing ones, that might motivate directionality. Work on psycholinguistics (Wasow 2002, 57) and grammaticalization (Hopper and Traugott 2003, 44) provides support for an account that models the language acquirer as both an active producer and passive interpreter of language data. Briscoe's (1998) evolutionary model of language change incorporates other factors beyond processing to explain change. Briscoe models linguistic selection in terms of the conflicting pressures of learnability, interpretability, and expressivity. Simplifying a bit, learnability is defined as the number of parameters that need to be set to acquire a target grammar, interpretability is parsing cost measured in terms of working memory load, and expressivity is defined as the cost of using a proper subset of one of the full languages defined by the grammar space. These pressures conflict over time; e.g., less expressive languages will have a lower average working memory load cost and fewer parameters to learn. Pairing a model of evolutionary language change like Briscoe's with a model of linguistic competence like that assumed in this thesis might lead to interesting insights about directionality, while perhaps reconciling the usage-based and acquisition-based accounts of directionality.

Style, genre, and change

As discussed in detail in Section 5.3, the model of linguistic competence discussed in this thesis assumes that linguistic competence accommodates and generates variation, and includes a quantitative, noncategorical, and nondeterministic component. Chapter 2 was devoted to exploring the formal foundations of StOT-LFG. Chapters 3 and 4 explored the empirical consequences of a theory of clause structure embedded in that framework. Throughout, the focus was on variation at the level of the individual text. A remaining gap in this account is an account of how style and genre (appropriately defined) fit into the model of change and variation examined in this thesis.

One possible solution was sketched in Section 5.3.3. In this approach, extralinguistic factors systematically boost or depress the ranking of specific constraints or groups of constraints. Some reasons why that particular approach might be too simplistic were discussed in Section 5.3.4. For example, the model should allow for constraints to index multiple categories or pragmatic meanings, as well as allow sets of constraints to cluster together to form certain styles. While formal probabilistic models of change and variation like StOT-LFG hold the promise that they will eventually dovetail with sociolinguistic work on stylistic variation, there is much more empirical and modeling work to be done on stylistic variation.

A related question is how to model the development of stylistic and genre variation over time. The work of Biber and Finegan (e.g., Biber and Finegan 1989) provides evidence that certain written genres in English such as letters drifted from written to more oral styles. These changes are indicated by changes in the sets of linguistic variables associated with these genres; e.g., in letters, the frequency of use of first- and second-person pronouns, contractions, and WH-questions — all associated with a more oral style — increases between the 19th and 20th century. Case studies of the sort described in Biber and Finegan's work pose challenging modeling problems for probabilistic models of change like that described in this thesis.