

Chapter 1

Introduction

Contents

1.1	Change and variation	2
1.2	The grammatical model	6
1.3	Case studies	8
1.4	Methodological issues	10
1.4.1	Dialect and contact issues	10
1.4.2	Genre issues	11
1.4.3	Corpus issues	11
1.5	Overview of the thesis	12

The central theoretical claim of this dissertation is that the appropriate grammatical model for syntactic change is one that incorporates a quantitative, noncategorical, and nondeterministic component. Building on work of Boersma (1998), Boersma and Hayes (2001), and Bresnan et al. (2001), I propose a novel formal framework for modeling syntactic change, Stochastic Optimality-Theoretic Lexical-Functional Grammar (StOT-LFG), that is firmly embedded in the inherent variability tradition (Weinreich et al. 1968; Bender 2001). I demonstrate that this model makes empirical and conceptual gains over previous models of syntactic change. This introduction lays out some starting assumptions about the nature of syntactic variation and change, as well as the methodological approach of the thesis.

1.1 Change and variation

Before I turn to some background assumptions about change and variation, I must define some terminology. *Innovation* is a difference or set of differences between the linguistic system (or systems) generating the primary linguistic data used by a language acquirer and the linguistic system constructed by the acquirer (Hale 1998, 2-3; Hale 2003, 345). *Spread* is the diffusion of a change from the innovator to those with whom the innovator comes in contact (Hale 1998, 5). *Change* is innovation plus spread.

Change always passes through an intermediate stage where the old coexists with the new, at the level of the community and for individuals (Harris and Campbell 1995, 48-49; Hopper and Traugott 2003, 49; Croft 2000, 49-50). That is, if β replaces α , then there must be an intervening stage in which α and β are in variation.¹ This is schematized in (1.1).

$$(1.1) \quad \alpha > \begin{matrix} \alpha \\ \beta \end{matrix} > \beta$$

The frequency of use of an innovated variant at the level of the speech community and the level of the individual often rises gradually, over the span of decades or centuries (Hiltunen 1983, 145; Kroch 1989b; Kroch 2001; Zuraw 2003).² This can be observed in the changing rate of use of an innovated variant by different generations of individuals in real time, as is amply illustrated by intratextual variability in the historical record; see, e.g., Pintzuk (1999).

Examples of gradual change can be observed in all areas of linguistics; e.g., phonological and syntactic change. For example, the raising and tensing of short *a* has been taking place in a number of different English-speaking communities for decades (Labov 1994; Kiparsky 1989; Guy 2003, 370).³ For syntax, the shift from OV to VO order in early English took place over a span of hundreds of years with examples of OV persisting through (at

¹This is “strategic gradualness” in the sense of Bennett (1981, 119) where “there is no sudden switch from strategy *x* to strategy *y*, but instead there is an intermediate stage where a speaker uses both *x* and *y*, in whatever proportions”, where a strategy is “a way of expressing a particular meaning”.

²Quantitative increase in use of a variant at the level of the *speech community* is equivalent to Trask’s (1996, 295) notion of “social gradualness”: “an innovating form is at first used by only a few individuals; over time, the innovating form comes to be used by even more individuals, while the conservative form is used by correspondingly fewer people, until the innovating form is (perhaps) the only one used in the community”. The variationist sense of gradualness that I am adopting here incorporates Trask’s notion of social gradualness but additionally claims that changes in the rate of use can be observed by comparing the variable language use of different generations of individuals in real time.

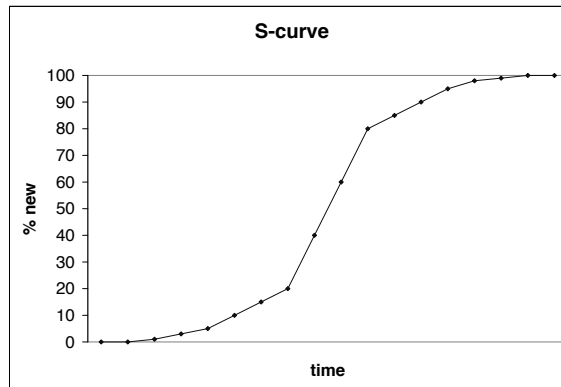
³For phonological change, it is important to distinguish a gradual shift in pronunciation from the gradual spread of a shift in pronunciation through the lexicon (i.e., lexical diffusion).

least) Middle English (Allen 2000; Kroch and Taylor 2001; Koopman and van der Wurff 2000; Pintzuk 1999; van der Wurff 1997). At each stage of the change, intratextual variability is observed, with individual writers often deploying both the OV and VO variants.

Not all variation leads to change. It is possible to find cases of long-term stable variation in all areas of linguistics. For example, in phonology, variation between the casual nonstandard suffix *-in* (as in *runnin'*) and the noncasual standard suffix *-ing* (as in, *running*) dates back to “at least the beginning of the 19th century and probably to the beginning of the 17th” (Labov 2001, 90). For syntax, persistent stable variation is also not hard to find. Labov (2001, 92) gives the example of negative concord, which he argues appears to be coextensive in time with *-ing/-in* as a sociolinguistic marker. Traugott (1992, 228) (see also Jespersen 1954, 81,132-153) provides evidence that variation between relative clauses with a relativizer (*that, who, and which*; e.g., *this is the cake **that I made***) and bare relatives (e.g., *this is the cake **I made***) — *contact clauses* in Jespersen’s terminology — goes back to Old English.

Languages often display uneven development, where one variant replaces another in a short period of time. Abrupt change often has its source in the external history of the language; e.g., population dislocations, migrations, invasions, conquests, etc. (Labov 1994, 24). For example, contact with Scandinavian has been argued to have had a dramatic influence on the history of the Northern dialects of early English (Geipel 1971; Jensen 1975; Thomason and Kaufman 1988; Kroch and Taylor 1997; Trips 2002; Fuss and Trips 2002; McWhorter 2002). More generally, it is assumed that linguistic change typically follows the logistic progression (the ‘S-curve’) illustrated by Figure 1.1 (Aitchison 1981, 97-107; Altmann et al. 1983; Croft 2000, 61-62, 183-190; Denison 2003; Devitt 1989, 34-48 Labov 1994, 65-67, Labov 2001, 449-454; Kroch 1989b; Osgood and Sebeok 1954, 155; Wang and Cheng 1970; Weinreich et al. 1968, 113). That is, 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).

Figure 1.1: The idealized timecourse of linguistic change



In historical studies, the vertical axis of graphs like Figure 1.1 have been identified with different aspects of linguistic space (Denison 2003, 60). For example, they have been used to represent lexical diffusion; e.g., sound change progressing word by word through the (relevant lexemes of the) lexicon (Aitchison 1981, 98-99; Wang and Cheng 1970). In the case studies presented in Chapters 3 and 4 of this thesis, the vertical axis represents the percentage total usage of an innovated variant, as in much other work in the variationist tradition. For many analysts, the different points in the trajectory of the idealized change in Figure 1.1 represents the composite usage of many individual writers/speakers. In the case studies presented in this thesis, I use individual texts from early English to demonstrate the (often monotonic) rise in frequency of an innovated variant (e.g., VO).⁴

Given these initial observations about the nature of linguistic change, we need a grammatical model that enables us to account for both the “steady quantitative rise of the rate of use of an innovation” (Guy 2003, 383), as well as long periods of stable variation. The observation that language change is always gradual — i.e., there is always a period where the innovated form and the old form coexist at the level of the community and the individual — suggests that the grammatical model for language change needs to treat variation not

⁴It is important to point out that the S-curve is not particular to language change. Rather, both linguistic and non-linguistic phenomena that exhibit changing percentages over time tend to follow an S-curve (Bennett 1981, 131; Kroch 1989b, 203-204; Lightfoot 1999, 102-104); e.g., epidemics.

merely as a “waystation on the road from one categorical state of the grammar to another” but as an “inherent feature of linguistic structure” (Guy 2003, 371).

Despite this, there has historically been a tension between, on the one hand, the apparatus that have been posited to model change (e.g., parameter resetting (Lightfoot 1991) and constraint reranking (Vincent 2000)) and, on the other hand, the observation that change does not proceed by discrete steps (Guy 2003; McMahon 2000, 105). For example, in parameter resetting and constraint reranking models of change, language change is treated as a catastrophic discrepancy between different grammars across real time. In this type of approach, the steady, quantitative rise of an innovative form must be dealt with by mechanisms external to the grammar; e.g., adaptive rules (Andersen 1973; Haspelmath 1999, 584).

Simplifying a bit, there are basically two traditions that attempt to resolve the tension between the grammatical models that have been used to explain change and the observation that change depends on variation. In the first tradition, the *competing grammars* tradition (Fries and Pike 1949; Kroch 1989b; Yang 2000), for each case of variable output that reflects opposed settings for a syntactic parameter, a pair of grammars must be postulated. In the second tradition, the *inherent variability tradition* (Weinreich et al. 1968; Bender 2001), knowledge of language accommodates and generates variation, and grammars include a quantitative, noncategorical, and nondeterministic component.

This thesis challenges the claim that syntactic change should be modeled in terms of adjustments in the probability distribution of competing grammars, arguing in favor of a particular inherent variability approach to syntactic change. Building on recent work in the framework of Stochastic Optimality Theory (Boersma 1998; Boersma and Hayes 2001; Bresnan et al. 2001), I develop in detail a grammatical model for syntactic change, which is an instance of the inherent variability tradition. This tradition also includes the variable rules approach (see references in Paolillo 2002), other optimality-theoretic approaches (partial ordering (Anttila and Fong 2000), floating constraints (Nagy and Reynolds 1997)), extensions of HPSG (Bender 2001), and extensions of the Principle and Parameters framework (Yang 2003). In the approach motivated in this thesis, grammars are conceived of as constraint rankings on a continuous scale of real numbers. Further, in the process of speaking or hearing (i.e., at any evaluation point) the rank of each constraint is slightly perturbed by adding a random value drawn from a normal distribution. This perturbation of constraint ranking leads to intraspeaker variation and, consequently, defines a probability

distribution over outputs from the grammar for a given input, where, roughly speaking, the input is the meaning of an utterance or the speech stream. Given this picture of linguistic competence, this thesis argues that syntactic change should be modeled as gradual movement in constraint strength along a continuous scale of real numbers.

The arguments that favor this type of model are of two types, conceptual and empirical. Chapter 2 and the conclusion address the former. Chapters 3 and 4 are focused primarily on the latter, showing that the StOT-LFG approach makes empirical gains over previous approaches.

1.2 The grammatical model

On my way to supporting the core theoretical claim of this thesis — i.e., that the appropriate grammatical model for syntactic change is one that incorporates a quantitative, non-categorical, and nondeterministic component — I demonstrate three things, each of which can be cleanly separated from the rest. First, I show that the type of constraint interaction available in Optimality Theory (Prince and Smolensky 1993) provides a more comprehensive and explanatory account of word order change and variation than is possible in parameter-based accounts (Kroch and Taylor 2001; Pintzuk 1999). Second, the framework of Stochastic Optimality Theory (Boersma 1998), a simple extension of standard Optimality Theory, is shown to satisfy the demand of providing accurate diachronic trajectories for change (Lightfoot 1999, 102-103). Third, it is shown that the particular framework advocated here, StOT-LFG, is the right sort of model for capturing syntactic variation and change.

Chapter 2 develops from the ground up the formal framework of StOT-LFG. Building upon work on the formal foundations of LFG (Kaplan and Bresnan 1995; Kuhn 2003b), as well as recent work on model-theoretic syntax and phonology (Blackburn and Gardent 1995; Blackburn and Meyer-Viol 1997; Pullum and Scholz 2001; Potts and Pullum 2003), an extensible description language, L^{lfg} , is defined that allows one to state both infeasible and defeasible constraints on LFG models. The framework and the theory embedded in that framework are stated at an unusual level of precision given current practice in diachronic generative syntax. I will now briefly explain my choice to delve so deeply into the formal foundations of the grammatical model advocated for here.

Above all, I am seeking to provide good underpinnings for the material on change, and

therefore am attempting for the first time to lay out all the details of the model, although only parts of the framework are relevant for the case studies in Chapters 3 and 4. Arguments for giving precisely constructed theories of linguistic structure include the following quote from Chomsky (1957, 5):⁵

Precisely constructed models for linguistic structure can play an important role, both negative and positive, in the process of discovery itself. By pushing a precise but inadequate formulation to an unacceptable conclusion, we can often expose the exact source of this inadequacy and, consequently, gain a deeper understanding of the linguistic data. More positively, a formalized theory may automatically provide solutions for many problems other than those for which it was explicitly designed. Obscure and intuition-bound notions can neither lead to absurd conclusions nor provide new and correct ones, and hence they fail to be useful in two important respects. I think that some of those linguists who have questioned the value of precise and technical development of linguistic theory may have failed to recognize the productive potential in the method of rigorously stating a proposed theory and applying it strictly to linguistic material with no attempt to avoid unacceptable conclusions by *ad hoc* adjustments or loose formulation.

Precisely formulated theories enable us to understand the domain of investigation more clearly. We can answer questions like *What are the models of this theory?* and *What are the consequences of this theory?*

There is a dearth of work on the formal foundations of Optimality Theory in comparison to other constraint-based approaches. Recent investigations include Kuhn (2003b) Potts and Pullum (2003), and Prince (2002). Eisner (1999) gives a handful of reasons to try to formalize optimality-theoretic constraints; see also Potts and Pullum (2003) and Asudeh (2001b). These include:

⁵See also Pollard and Sag (1994, 6-14) for the type of methodology that I am adopting in this thesis.

- (i) “Results in an explicit, falsifiable theory of UG”
- (ii) “Simplifies that theory, exposing formal similarities among constraints”
- (iii) “Aids descriptive work by providing well-motivated and well-formalized constraints and representations”

The theory presented in this thesis, like other work in optimality-theoretic syntax, is complicated and often the empirical consequences of the interaction of different constraints are not straightforward. For this reason, there is a necessity for unambiguous definitions of constraints. The empirical hypotheses embodied by the individual constraints are interpreted as a set of soft (violable) and hard (inviolable) constraints in the description language L^{fig} . It is demonstrated that the description language is extensible to other phenomena not investigated in detail here. Each constraint is also expressed informally in English.

The particular formalism that I have employed to represent syntactic structures is well-suited as a representational formalism for optimality-theoretic syntax. Kuhn (2003b, 69ff.) (see also Johnson 2002) shows that unrestricted optimality-theoretic systems lead to problems with decidability for processing tasks (parsing/recognition and generation). In a mathematically well-studied formal architecture like that of Lexical-Functional Grammar, though, there is a candidate generation algorithm that permits enumeration of OT candidate sets and computationally tractable evaluations of the optimality of candidates (Kuhn 2003b, 74-75).

By laying out in detail the mathematic underpinnings of StOT-LFG, this work contributes to the growing body of literature on the formal foundations of Optimality Theory, as well as contributing to work in LFG; e.g., by unifying the specification of c-structure displayed in early LFG work (Kaplan and Bresnan 1995) with recent optimality-theoretic approaches (Sells 2001c).

1.3 Case studies

The empirical core of this dissertation focuses on two case studies of change and variation in early English. The phenomena discussed in Chapters 3 and 4 have been used to

motivate the competing grammars approach. Chapter 3 serves as an extended demonstration of how StOT-LFG can be applied to syntactic change, as well as a concentrated empirical argument against one particular variant of the competing grammars approach. In that chapter I examine the decline of right-headedness (e.g., OV to VO) in Old and Middle English. Drawing on data from secondary sources, as well as novel data from the York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE; Taylor et al. 2003) and the Penn-Helsinki Parsed Corpus of Middle English, Second Edition (PPCME2; Kroch and Taylor 2001), it is demonstrated that a StOT-LFG account makes empirical gains over the competing grammars account; e.g., deriving the absence of the reverse brace construction (as in, *you keep God's commandments will*). Chapter 4 studies variation and change in the syntax of subjects in early English, drawing on secondary sources as well as new evidence from the YCOE and the PPCME2. The StOT-LFG approach is shown to capture with precision the gradual shift in the status of subject pronouns from nonprojecting (weak) to projecting (strong), as well as providing an account of intraspeaker variation in the position of full NP subjects. Further, I show that the StOT-LFG approach is able to account for the *stochastic generalization for change*: factors that influence the choice of variants at one stage of a change can often be found to have a categorical influence at another stage.

The two problems I tackle in this work are very well-known puzzles in historical English syntax. That being the case, my individual examples are of necessity often drawn from secondary sources. Wherever possible, though, I have supplied new evidence from the parsed corpora mentioned above, as well as making heavy use of the Old English Corpus⁶ and the Middle English Dictionary.⁷ The system of reference used for individual examples from Old English and Middle English is that of the source that they were drawn from.

Before I turn to some methodological issues, I must briefly note some of the simplifying assumptions about periodization that surface in Chapters 3 and 4. I make use of various partitions of the history of English like “Old English” and “Middle English”, not to mention more refined periodizations like “late Old English” and “early Middle English”. The periodization I assume is in (1.2), while also agreeing with Fisiak (1994, 54) that precise dates like 1066 (as the beginning of Middle English) are linguistically arbitrary:

⁶<http://ets.umdl.umich.edu/o/oec/>

⁷<http://ets.umdl.umich.edu/m/med/>

- (1.2) **Early Old English:** 700 — 925
Late Old English: 925 — 1150
Early Middle English: 1150 — 1325
Late Middle English: 1325 — 1500

The periodization in (1.2) is not uncontroversial, but will do for present purposes. Nothing in the analyses in Chapters 3 and 4 depends on this particular periodization. Recent work by Blake (1994), Fisiak (1994), and Lass (2000) problematizes the traditional system of periodization of early English. I hope to return to this issue in future work.

1.4 Methodological issues

1.4.1 Dialect and contact issues

As discussed above, I illustrate the trajectory of syntactic change in Chapters 3 and 4 by comparing the frequency distribution of variants for different generations of writers (or speakers). There is a serious risk in comparing Old and Middle English texts without taking dialect or language contact into account. Late Old English texts are mainly West-Saxon (southerly), while early Middle English texts like the *Peterborough Chronicle* are of the East Midlands dialect (Peterborough, Northamptonshire; northern border of the South Midlands area). If we compare texts like the *Peterborough Chronicle* “with, say, earlier parts of the *Anglo-Saxon Chronicle* we are really comparing apples and pears” (Allen 2000, 6). Contact with Scandinavian, for example, may have accelerated some of the changes discussed in this thesis, at least in the more northerly dialect areas (Kroch and Taylor 1997; Trips 2002; Fuss and Trips 2002; McWhorter 2002). For instance, it could be the case that contact with Scandinavian had a dramatic influence on the syntax of writers from the Peterborough area, causing an abrupt change in the ambient data for language learners (Allen 2000, 7). For these reasons, I have tried throughout this thesis to compare texts from roughly the same dialect area, namely southerly texts.

1.4.2 Genre issues

Similar issues arise when comparing texts from different genres. For example, in work based on the *Helsinki Corpus of English Texts*, it has been demonstrated that linguistic innovations take place earlier in more speechlike genres like personal letters and trial proceedings than written genres like legal documents (Nevalainen and Raumolin-Brunberg 2003, 2,192). In recent work, Warner (2000) demonstrates that genre (i.e., more oral vs. less oral texts) has a significant impact on the distribution of periphrastic *do*. Further, individual writers can vary their style depending on the genre (Nevalainen and Raumolin-Brunberg 2003, 192). For these reasons, when comparing frequency distributions from different texts in Chapters 3 and 4 I utilize evidence from the same genre (e.g., religious texts) where possible. In the conclusion, I discuss how the framework advocated for in this thesis might be extended to account for the interaction of extralinguistic factors and linguistic variables.

1.4.3 Corpus issues

Electronic corpora like the YCOE and PPCME2 have made an invaluable contribution to research on historical morphosyntax, enabling researchers to analyze large bodies of syntactically parsed text. As noted above, this thesis makes heavy use of a variety of electronic corpora. For this reason, it is necessary to spend some time highlighting several of the issues surrounding the empirical base of this thesis.

First, the date of a manuscript must be distinguished from the (likely) date of composition of the text the manuscript contains. Later scribes may have altered the original text, substituting newer syntax for the older syntax of the original text.⁸ For this reason, researchers cannot simply assume that the date of an example is the same as the date of composition. This is especially true if there is a large gap between the date the text was composed and the date of the manuscript.

Second, electronic corpora add a new layer of issues beyond possible discrepancies between the original text and the edited manuscript that contains that text.⁹ It is rarely the case that it is clear from text drawn from electronic corpora what emendations were made by the manuscript editor. Likewise, it is not typically possible to figure out what the manuscript editor's policy on punctuation was. This is the case for the corpora used in this

⁸See, for example, Allen (1992, 5-6) on the syntax of the verb *behofian* 'behoove' in Old English.

⁹This discussion draws upon van Bergen (2003, 13-17).

thesis. Further, it is inevitable that electronic corpora contain typographical errors. Ideally all the examples drawn from corpora should be verified against the original manuscripts. However, this is not always practical for large numbers of examples. For the case studies in Chapters 3 and 4, I verified most of the examples drawn from electronic corpora against the edited manuscripts that were their source.

1.5 Overview of the thesis

This thesis contributes both an empirically verifiable theory of phrase structure and, for the first time, a comprehensive and extensible formal framework for modeling syntactic variation and change in Stochastic Optimality Theory. Apart from the introduction, the thesis is divided into four chapters. The first chapter lays out the formal framework, StOT-LFG, that will serve as the basis for the two empirical chapters. The second chapter discusses the decline of right-headed clauses in English; e.g., the OV to VO shift. The third chapter analyzes change and variation in the syntax of early English full noun phrase subjects and subject pronouns. The final chapter serves both as a summary of what has been accomplished in the thesis and as an apologia for the inherent variability tradition, as well as charting a course for future work.