Grammatical Constraints on Variation:
'Be' in the *Survey of English Dialects* and (Stochastic) Optimality Theory[1]

Joan Bresnan and Ashwini Deo
Stanford University
Draft of May 7, 2001

A single speaker's production of variable outputs for the same input often reflects categorical generalizations across grammars. Variation in subject-verb agreement in English dialects and varieties has just this property. For example, previous studies of verb agreement in 'nonstandard' English varieties have observed a reduction of variation with plural (vs. singular) subjects or negative (vs. affirmative) sentences; these are typologically marked grammatical contexts in which contrasts are often categorically neutralized across languages (Ihalainen 1991, Cheshire 1991, 1996, Cheshire, Edwards, and Whittle 1989, 1993, Schilling-Estes and Wolfram 1994, Anderwald 1999, forthcoming). In the present study, the grammatical structuring in individual patterns of variable subject-verb agreement with *be* extracted from the *Survey of English Dialects* (Orton et al. 1962–71) bears striking relations to categorical varieties, and both can be explained in a unified way within Optimality Theory with stochastic evaluation (Boersma 1998). The stochastic grammars used here have interesting implications for conceptions of linguistic competence, variation, and historical change.

1. Inventories from the *Survey of English Dialects*

Although *be* variation is the subject of much important work on all varieties of English, England may have the widest variety of *be* inventories (Schilling-Estes and Wolfram 1994: 277), and this was our reason for studying *be* in the *Survey of English Dialects* (*SED*, Orton et al. 1962–71).

We constructed inventories of present tense *be* forms by extracting questionnaire responses by individual speakers from the *SED* and organizing them paradigmatically. 53% (= 161) of the total possible inventories were extracted. For each questionnaire item the *SED* groups all of the responses within a specific county, but it also indexes individual respondents across questions. To construct the *be* inventories, we grouped the responses to all *be* questions for each individual respondent. Many responses give

[1]We wish to thank Brady Clark, Andrew Garrett, Jane Grimshaw, Dylan Herrick, Junko Hibiya, Jonas Kuhn, Wouter Kusters, Hanjung Lee, Roger Levy, Chris Manning, Line Mikkelsen, Peter Sells, Devyani Sharma, and Elizabeth Traugott for their comments on earlier versions of this work. We remain solely responsible for all errors of fact and interpretation. This work is based in part on work supported by the National Science Foundation under Grant No. BCS-9818077.

varying forms to the same or similar questionnaire items, and these are our sources for the variable inventories. Each inventory thus represents the grammar of an individual speaker or a group of speakers from whom the same input/output pairs were elicited.

Because of the limited scope of our initial constraint set (based on Grimshaw 1997b, in press, and Bresnan 2001a, in press) and the limitations of the data, we decided to restrict the present study to present-tense non-negated *be* inventories in basic, unmarked sentences, excluding those inventories that can only be characterized in terms of positional variation from these basic contexts. These limitations are discussed in the Appendix. 39% (= 63) of the inventories extracted are included by this criterion.

In this section we survey the inventories we extracted. We collect together all inventories that vary only in pronunciation by giving orthographic representations of the original IPA transcriptions. The numbers following county names are the labels for individual respondent questionnaires in *SED*.

1.1. Inter-Speaker Variation

Figure 1 shows the patterns of inter-speaker variation in our data.

Note that the 'Northumberland' pattern (4) is the same pattern we see in the Standard past tense *be*: *I was, you were, she was, we/you/they were.* Note also that the inventory in (5) shows a complete loss of all agreement contrasts. Parallel inventories based on *am*, *are*, *is* have also been reported, although we did not find these in our data: *I/you/she/we/you/they am here*, *I/you/she/we/you/they are here*, *I/you/she/we/you/they is here* (Trudgill 1990: 98). Past tense in West and East Midlands shows a similar loss of all agreement contrasts (Cheshire, Edwards, and Whittle, 1993: 80): *I were singing. So were John. Mary weren't singing.*

Overall, these present tense *be* inventories can be characterized as follows:

- There are from zero to three person distinctions in the singular, with absence of all person distinctions in the plural.
- The absent distinctions in the singular include first person, second person, first and second person, first and third, and all three.[2]
- The morphs used for the generalized forms vary among *are*, *be*, and *is*. It is striking that the same abstract patterns of contrasts are sometimes instantiated with different morphs; we can thus distinguish the inventory of specific forms from the inventory of abstract contrasts expressed for subject verb agreement. It is the latter that interest us here.

To reemphasize the last point, we note particularly that the problem of accounting for the occurrence of individual morphs—the historical suppletion problem—is *not*

[2] Additional possibilities occur among the excluded inventories, so we do not assume these to be exhaustive.

Figure 1: Patterns of Inter-Speaker Variation

(1) Yorkshire

	SG	PL
1	is	are
2	is	(are)
3	is	are

(2) Kent

	SG	PL
1	are	are
2	are	(are)
3	is	are

(3) Somerset

	sg	pl
1	be	be
2	be	(be)
3	is	be

(4) Northumberland

	sg	pl
1	is	are
2	are	(are)
3	is	are

(5) Sussex

	sg	pl
1	be	be
2	be	(be)
3	be	be

(6) Suffolk/Standard

	sg	pl
1	am	are
2	are	(are)
3	is	are

(7) Derbyshire

	sg	pl
1	am	are
2	art	(are)
3	is	are

(8a) Dorset

	sg	pl
1	be	be
2	art	(be)
3	is	be

(8b) Hampshire

	sg	pl
1	be	be
2	bees(t)	(be)
3	is	be

(9) Berkshire

	sg	pl
1	be	be
2	beest	(be)
3	be	be

(1): Yorkshire 7, Lancashire 1, Westmoreland 3, Northumberland 9
(2): Kent 7
(3): Somerset 6, Hampshire 2,5
(4): Northumberland 6
(5): Sussex 2,5
(6): Suffolk 1,2,3,4,5, Rutland 1,2, Huntingdonshire 1,2, Middlesex & London 1,2, Staffordshire 5,10, Oxfordshire 4, Monmouthshire 6, Kent 5, Surrey 1, Northumberland 1,3, Lancashire 6,10,11,13,14, Man 1,2
(7): Derbyshire 3,4, Cornwall 5,7, Lancashire 9
(8a): Dorset 3, Cornwall 1
(8b): Hampshire 6, Gloucestershire 3,4
(9): Berkshire 2

Note: Inventories where the phonetic transcription of plural forms includes *wə, ðə* (e.g. (1)) are analyzed as phonetic variants of *we're, they're.*

our concern in the present study. Our focus rather is on the patterns of person/number contrasts that are revealed by the inventories. These abstract patterns represent variation in subject-verb agreement contrasts, the central topic of our inquiry.

1.2. Intra-Speaker Variation

The patterns of intra-speaker variation (variable inventories) are shown in Figures 2 and 4. These were extracted from the questionnaires for present tense *be* by the same criteria we outlined above, which exclude inventories characterizable only by means of positional variation.

Figure 2: Patterns of Intra-Speaker Variation

(10) Kent

	sg	pl
1	am, are	are
2	are	(are)
3	is	are

(11) Yorkshire

	sg	pl
1	am	are
2	art, are	(are)
3	is	are

(12) Berkshire

	sg	pl
1	be	be
2	be	(be)
3	is, be	be

(13) Gloucestershire

	sg	pl
1	be	be
2	beest	(be)
3	is, be	be

(14) Monmouthshire

	sg	pl
1	be, am	be
2	beest	(be)
3	is	be

(15) Northumberland

	sg	pl
1	is, am	are
2	are	(are)
3	is	are

(16) Berkshire

	sg	pl
1	be	be
2	be, are	(be)
3	be, is	be

(10): Kent 3, Bedfordshire 1,3
(11): Yorkshire 21
(12): Berkshire 3, Oxfordshire 3, Sussex 6
(13): Gloucestershire 3
(14): Monmouthshire 1, Gloucestershire 7
(15): Northumberland 8
(16): Berkshire 4

The inventories of Figure 2 all have simple relations to the categorical inventories of Figure 1. As illustrated in Figure 3, (10) can be analyzed as combining the Kent pattern in Figure 1 (2) with the “Standard” pattern in Figure 1 (6). Likewise, (11), viewed as an abstract pattern of constrasts, combines the Derbyshire/Cornwall pattern in Figure 1 (7) with the Standard pattern in Figure 1 (6). The same holds for all of the variable inventories of Figure 2. Viewed as abstract patterns of contrast, each of the variable inventories is a combination of the categorical patterns.

Note that ‘Monmouthshire’ (14) can be viewed as combining Hampshire/Gloucestershire (8b) with the abstract pattern of contrasts of Derbyshire/Cornwall (7), having three person distinctions in the singular and no person contrasts in the plural.

Figure 3: Variable Combinations of Categorical Patterns

(10) Kent = (6) Suffolk/Standard + (2) Kent

	sg	pl
1	am, are	are
2	are	(are)
3	is	are

	sg	pl
1	am	are
2	are	(are)
3	is	are

	SG	PL
1	are	are
2	are	(are)
3	is	are

(11) Yorkshire = (6) Suffolk/Standard + (7) Derbyshire

	sg	pl
1	am	are
2	art, are	(are)
3	is	are

	sg	pl
1	am	are
2	are	(are)
3	is	are

	sg	pl
1	am	are
2	art	(are)
3	is	are

Also, (12) = (5) + (3), (13) = (9) + (8b), (14) = (8b) + (7), (15) = (6) + (4), (16) = (5) + (8a).

Also note that *are* in the context of (16) appears to be a [2SG] form parallel to *art* in Figure 1 (8a).

Residual Inventories The three remaining variable inventories that met our criteria for selection are shown in Figure 4. Note that for these we do not make any assumptions about what fills the missing [2PL] cell, because there is variation in the plural. We discuss these further below. ((1a) and (1b) are grouped together because they have *am* (contracted as in *we'm, they'm*) as a variant plural form.)

Figure 4: Residual Variable Inventories

(17) Durham 6

	sg	pl
1	is	are
2	is	
3	is	are, is

(18a) Surrey 4

	sg	pl
1	am, are	am, are
2	are	
3	is	am, are

(18b) Kent 1

	sg	pl
1	am, are	are
2	are	
3	is	am, are

We can draw several conclusions from this survey of variable *be* inventories:

- Most of the variable inventories appear to be combinations of two invariable inventories. To a substantial degree, the variation of inventories within grammars reflects variation of inventories between grammars.

- In general, the variable inventories have fewer plural forms than singular forms.[3] This is partly true because the variable inventories, like the invariable inventories, lack person contrasts in the plural, so that in every inventory the total number of distinct forms in the plural is less than the total number of distinct singular forms. It is also true of most of the variable inventories that for each person (disregarding the second person, where data are incomplete), the number of variant forms in the singular is greater than the number of variant forms in the plural. Thus there is generally a reduction of variation as well as a reduction of contrast in the plural compared to the singular. In this sense, intra-speaker variation appears to be constrained by the same factors that play a role in inter-speaker variation.

- The general (or 'default') form is often in free variation with more specific forms, as in the inventories of Figure 2. This patterning is not naturally explained by a model using blocking or preemption of defaults by more specific forms, nor by an ordinal ranking of violable constraints (ordinal OT).

2. The Framework of Stochastic Optimality Theory

We now turn to the framework we use for formally analyzing the surveyed inventories. Optimality Theory with stochastic evaluation, which we call 'stochastic OT' or 'StOT', was originally developed by Paul Boersma as part of a theory of functional phonology that addresses the learning of categories, variation, optionality, and probability (1997, 1998, 2000, Boersma and Hayes 2001). It is one of a family of generalized OT frameworks that address variation (see Anttila 2000, Boersma 1999b, Hibiya 2000, and Boersma and Hayes 2001 for reviews). Stochastic OT is distinguished by a particularly well-developed underlying theory, including an associated Gradual Learning Algorithm (GLA), and an implementation within the freely available cross-platform Praat computer program (Boersma 1999a, Boersma and Weenink 2000)[4]

Related OT work on syntactic variation includes van Oostendorp 1997, Nagy and Heap 1998, and Anttila and Fong 2000. Related work using the stochastic OT framework in particular includes Asudeh 2001, Lee 1999, 2000, Koontz-Garboden 2001, Bresnan, Dingare, and Manning 2001, Dingare 2001, and Clark 2001.

2.1. Inputs and Outputs

An OT grammar can be viewed as a function from INPUTs to OUTPUTs. We take the morphosyntactic INPUT to be language-independent content drawn from the space of possible lexical and grammatical contrasts and the OUTPUT to consist of language-specific forms with varying analyses of that content. The relation between INPUT

[3]This is sometimes said to be a general property of Germanic, but in modern Icelandic, and in Old Icelandic as well to a lesser extent, in most paradigms there is only one person distinction in the singular—1st against 2nd and 3rd, or 1st and 3rd against 2nd person—while 1st, 2nd, and 3rd person are distinguished in the plural (Wouter Kusters, p.c., April 6, 2001).

[4]The GLA is also implemented in OTSoft, also freely available (Hayes, Tesar, and Zuraw 2000).

and OUTPUT is determined by the ranking of violable constraints hypothesized to be present in all grammars.[5]

The overall structure we assume for an OT grammar of English verb forms is shown in Figure 5 (Bresnan 2001a, in press a,b). The INPUT is represented here by a feature structure (an f-structure, or attribute-value matrix) while the OUTPUT is represented by pairings of expressions and feature-structures in correspondence.[6] This conception of INPUT and OUTPUT draws on a mathematically and empirically well-understood representational basis (see Bresnan 2001b, Kuhn 1999b, 2000, 2001 and references). Each candidate shown in Figure 5 consists of a word form such as *art* standing for a 'functional' verb (or auxiliary) and a candidate feature structure. Inputs are fully specified for person/number features. Candidate expressions for each INPUT are generated by GEN and evaluated according to an EVAL function which, given a set of ranked constraints, defines the OUTPUT to be the candidate which best satisfies the highest ranked constraint on which it differs from its competitors (Grimshaw 1997a.)

2.2. Stochastic Evaluation

Stochastic OT adds two essential ideas to this standard OT picture. The first is that constraints are ranked on a continuous scale of real numbers, rather than an ordinal scale. Hence, constraints differ in distance, as well as dominance. The second idea is that at each evaluation the rank of a constraint varies slightly. That is, the rank it has in the grammar is the mean of a normal distribution or 'bell curve' of variant values that it has when applied in evaluations, as illustrated in Figure 6.[7]

These two ideas are stated in (1):

(1) **Stochastic OT** (Boersma 1997, 1998, 1999a, Boersma and Hayes 2001):

(i) **ranking on a continuous scale:** constraints are ranked on a continuous scale of real numbers, rather than a discrete ordinal scale;

(ii) **stochastic evaluation**: at each evaluation the rank of each constraint is perturbed by temporarily adding to its ranking value a random value drawn from a normal distribution, with the result that the actual rankings that determine the winner vary for each production of an output.

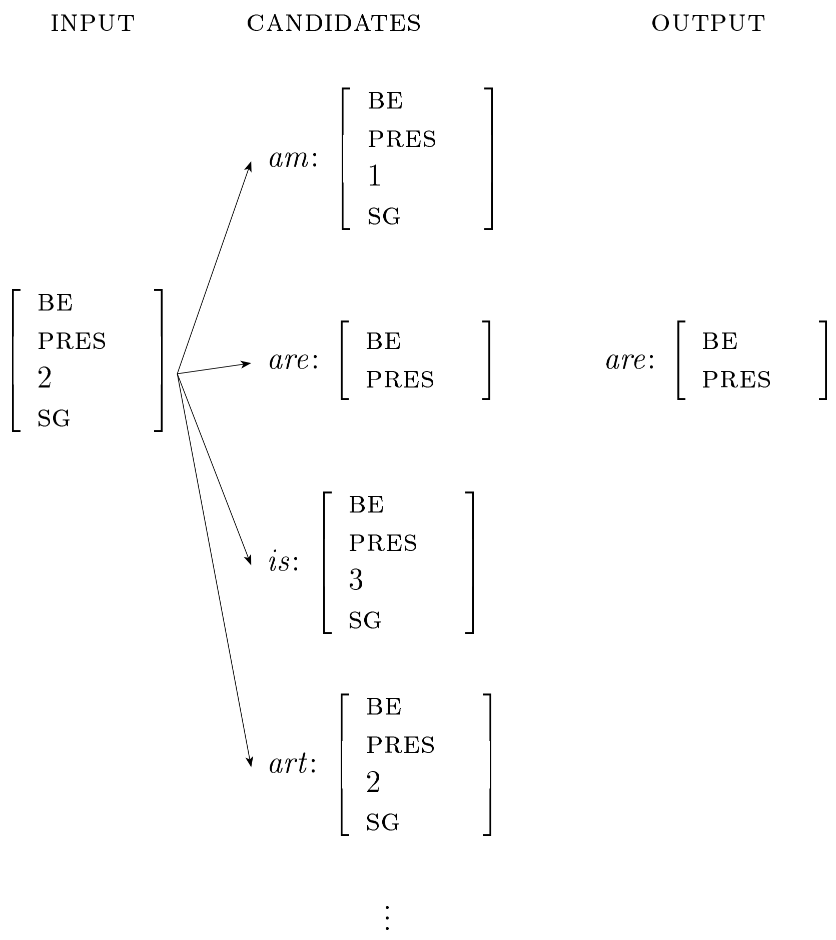
As explained by Boersma and Hayes (2001), a stochastic OT grammar can gen-

[5]Note that Stochastic OT as a framework does not require that constraints be universal and/or innate, and in fact Boersma's theory of functional phonology (1998) is a well articulated alternative.

[6]In Figure 5 PRES may be regarded as an abbreviation for [TENSE PRES], [2] for [PERS 2], [SG] for [NUM SG], and [BE] for [PRED 'BE'] in the customary attribute-value notation in which $+feature$ is rendered[$feature$ $+$] (Johnson, 1988). The expressions paired with each f-structure actually consist of an abstract characterization of word class properties, such as V^0 or I^0, and a language-particular pronunciation, such as *is*; we use the latter to abbreviate the entire expression.

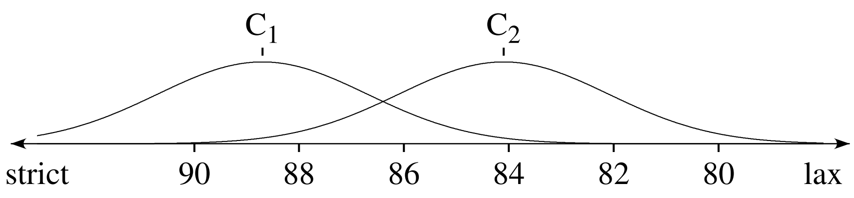
[7]The diagrams in Figures 6–9 are adapted from Boersma and Hayes (2001).

Figure 5: OT Grammar of English *be* Verb Forms



GEN: INPUT→ CANDIDATES
EVAL: CANDIDATES → OUTPUT

Figure 6: Relation of two constraints C_1, C_2 under stochastic evaluation:



erate both categorical and variable outputs. Categorical outputs arise when the constraints active in selecting them over their competitors are spread far apart on the continuous scale, so that the stochastic variation in ranking values has no discernable effect. In Figure 7, for example, $C_1 \gg C_2$ and the two constraints are spread far enough apart that the bulk of their ranges of variation (which is illustrated in a simplified way by the ovals) does not overlap. As the distance between constraints increases, interactions become vanishingly rare, reaching a point where variant outputs lie beneath any given error threshhold, or beyond the life expectancy of the speaker.

Figure 7: Categorical constraint ranking with ranges of variation:

Variable outputs arise when the active constraints are close enough together for the variation in their ranking values to interact with some observable frequency. This possibility is illustrated in Figure 8, where the bulk of the ranges of variation of two constraints overlaps. Here again $C_1 \gg C_2$, but with some discernable frequency during stochastic evaluation C_1 will be ranked at a point in its lower range, call it c_1, while C_2 is simultaneously ranked at a point c_2 in its higher range. As shown in Figure 9, C_2 will then temporarily dominate C_1 in selecting the optimal output, possibly producing a different output. The frequency of this reversal depends on the ranking distance between constraints and the standard deviation in ranking variance during evaluations (which is assumed to be the same across constraints). If we take the standard deviation to be zero, the constraints are always evaluated in the same strict domination sequence, and we have ordinal OT (Prince and Smolensky 1993). Stochastic OT is thus a generalization of ordinal OT. Its associated Gradual Learning Algorithm (GLA) can learn grammars robustly from variable data (Boersma 1997, 1998, 2000, Boersma and Hayes 2001).

Figure 8: Free constraint ranking with ranges of variation:

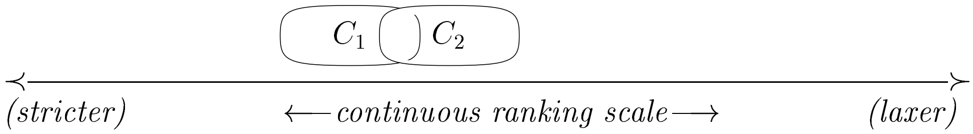
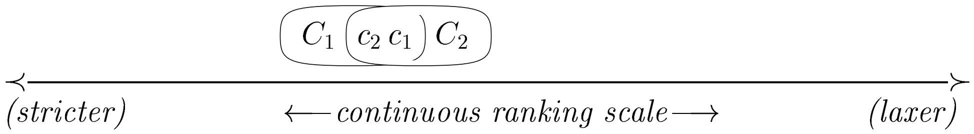


Figure 9: Reversal of constraint dominance:



3. A Theory of Variable *be* Agreement

The fact that person/number inventories can be derived in OT through constraint reranking is first established by the work of Grimshaw (1997b, in press) on Romance clitics. Our theory builds on this work and that of Bresnan (2001a, in press a).

3.1. The Constraint Set

There are three types of constraints: the faithfulness constraints Max and Ident, which compare a candidate to the input, the markedness constraints *Feature, which assess the well-formedness of the candidate in terms of its featural complexity; and a 'social deixis' constraint, to be discussed below. The faithfulness and markedness constraints are quite familiar types from research in OT phonology.[8] The basic intuition is that markedness constraints penalize complex or 'difficult' structures, and so tend to erode contrasts. Faithfulness constraints, in contrast, require that features of the input content be preserved in the output expression; they thus serve the communicative function of expressing contrasts in content, protecting content against the eroding effects of markedness constraints on forms. A particular language harmonizes these conflicting constraints by prioritizing (ranking) them.

The Max constraints penalize candidates which delete features of the input:[9]

(2) Max(num), Max(pers): If the input has a num (respectively pers) value, the candidate feature-structure has a num (respectively pers) value. Abbreviations: Max(n), Max(p) and mxn, mxp.

In (3) candidate 1 violates both Max(pers) and Max(num) because it lacks values for pers and num. Candidate 2 satisfies both Max constraints, even though the value of its pers feature differs from the input.

[8]The Dep class of constraints, which penalize the addition of features not in the input, are omitted to simplify the exposition because they do not distinguish among the candidates under discussion here.

[9]Of course, derivational operations of feature deletion and rewriting are not involved; rather, these are epiphenomenal consequences of the parallel optimization of candidates that may diverge from the given input in various ways.

(3) example input: $\begin{bmatrix} \textsc{num} & \textsc{sg} \\ \textsc{pers} & 2 \end{bmatrix}$

candidate 1: *be*: $\begin{bmatrix} \textsc{num} & \\ \textsc{pers} & \end{bmatrix}$

candidate 2: *is*: $\begin{bmatrix} \textsc{num} & \textsc{sg} \\ \textsc{pers} & 3 \end{bmatrix}$

The IDENT constaints penalize candidates which change the value of a feature present in the input to another value[10]

(4) IDENT(NUM), IDENT(PERS): If the input and candidate both have a NUM (respectively PERS) value, then the values are identical. Abbreviations: ID(N), ID(P) and idn, idp.

In (3), candidate 2 violates IDENT(PERS) but satisfies IDENT(NUM). Candidate 1 does not violate the IDENT constraints, because it has no correspondents of the input NUM or PERS feature values.

Different faithfulness constraints may be instantiated for various morphosyntactically defined domains; this is called 'positional faithfulness' in phonology (Urbanczyk 1995; Benua 1995). English has three inflectional classes for present-tense verbs (*be*, modal verbs, and other verbs), for which there are three families of separately rankable faithfulness constraints (Bresnan 2001a, in press a). We will be concerned here mainly with faithfulness in the domain of *be*. The above faithfulness constraints are thus implicitly indexed to this domain.

The markedness constraints in (5) penalize the featural complexity of candidates:[11]

(5) *1, *2, *3: Avoid a candidate if its PERS value is 1 (respectively 2,3).
*SG, *PL: Avoid a candidate if its NUM value is SG (respectively PL).

Finally we come to the "social deixis" constraint we have added to the constraint set of Bresnan (2001a, in press a):

(6) SOC: Avoid singular expressions for second person inputs. That is, mark a candidate if the input PERS value is 2 and the candidate NUM value is SG.

[10] See n. 9.

[11] Another markedness constraint enforces the fusion of PERS and NUM by requiring the same candidate to express both features simultaneously:

> FUSE-P/N: Avoid a candidate if it has a PERS value and no NUM value or a NUM value and no PERS value.

Violation of FUSE-P/N would allow splitting person exponence from number exponence, a typological possibility which English does not take. Without pursuing a detailed analysis, Bresnan (in press a: n. 6) assumes that some such constraint exists. It can be eliminated without distorting essential properties of our analysis, and we do so to keep the exposition as uncomplicated as possible.

The avoidance of too direct reference to the second person is a recurrent cross-linguistic phenomenon, with pragmatic and/or sociolinguistic motivations (Brown and Levinson 1978/1987) which may become formally crystallized in grammars. Constraint (6) is hypothesized as a formal reflex of this pressure. If it outranks faithfulness to number, it will require a nonsingular expression of the second person. To see how this constraint is interpreted, consider (3) again. Here the input is [2SG]. Candidate 2 is [3SG], but is still ruled out by (6). Candidate 1 is not marked by constraint (6).

The social deixis constraint is required to account for the Northumberland inventory in Figure 1 (4): *I is, you are, she is, we/you/they are*, as discussed below. A constraint marking [2SG] candidates on the basis only of the featural content they carry would cause *art* to be avoided, but could not prevent the use of *is* instead. Constraint (6) penalizes both *art* and *is as expressions of second person input.*[12]

3.2. Constraint Ranking and Variable Outputs

In the stochastic OT grammar the above constraints are ranked on a continuous scale and evaluated stochastically. One way to understand the model is to compare the output distributions of the same grammars under ordinal and stochastic evaluation. An OT grammar for Suffolk/Standard English is shown in the tableaux of Figure 10.

In the first tableau of Figure 10 the most faithful candidate *am* is the optimal output for [1SG] input. As in Bresnan (2001a, in press b) this result follows from the ranking of *3, *1, and *SG below the faithfulness constraints: the featural complexity of third singular or first singular forms like *am* is overridden by their faithfulness to input person and number. Similarly, *is* will be the output of [3SG] input.

The second and third tableaux of Figure 10 evaluate [2SG] and [3PL] inputs, respectively. The ranking of *2 and *PL above the MAX faithfulness constraints means that it is worse to carry these features than to be unfaithful to the input. For this reason a second person singular candidate (represented by *art* in the tableau) is not the optimal output for [2SG] input. The candidate *is*[3SG] nicely avoids this problem, but it differs in person from the input, violating highly ranked IDENT(PERS). An impersonal singular *is*[SG] avoids both of these problems, but it runs afoul of the social deixis constraint SOC. The plural candidates are too marked, by the high ranking of *PL, and so the only viable candidate is the general form *are*[ ]. The general form *are* is optimal under this ranking for [2SG] and all PL inputs (Bresnan 2001a, in press b).

These tableaux can be produced by both of the stochastic OT grammars shown in Figure 11. Visual inspection shows that the constraints of these grammars differ slightly in their ordinal rankings, both from each other and from the tableaux in Figure 10, but they fall into the same three strata consisting of MAX(PERS) ('maxp'), the constraints above it (*PL, *2, SOC, IDENT(PERS) and IDENT(NUM)), and those below it (*1, *3, *SG, MAX(NUM)). The constraints within each stratum are not crucially ordered for the Standard English categorical output, and under ordinal evaluation

[12] Smolensky (1998) also proposes constraints on input structures to account for gaps in the expression space in the domain of syntax.

Figure 10: Tableaux of a Grammar of Standard English

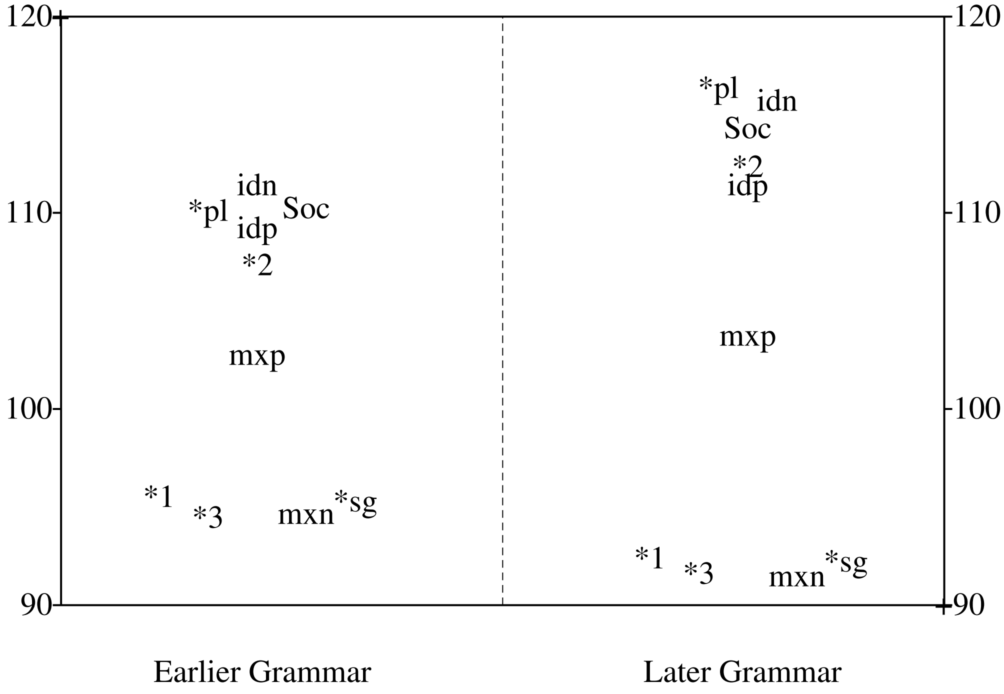
[1sg]	*PL	ID(N)	*SOC	ID(P)	*2	MAX(P)	*1	*3	*SG	MAX(N)
☞ am[1sg]							*		*	
art[2sg]				*!	*				*	
is[3sg]				*!				*	*	
is[sg]						*!			*	
are[pl]	*!	*				*				
are[]						*!				*
are[1pl]	*!	*					*			
are[2pl]	*!	*		*	*					
are[3pl]	*!	*		*				*		

[2sg]	*PL	ID(N)	*SOC	ID(P)	*2	MAX(P)	*1	*3	*SG	MAX(N)
am[1sg]			*!	*			*		*	
art[2sg]			*!		*				*	
is[3sg]			*!	*				*	*	
is[sg]			*!			*			*	
are[pl]	*!	*				*				
☞ are[]						*				*
are[1pl]	*!	*		*			*			
are[2pl]	*!	*			*					
are[3pl]	*!	*		*				*		

[3pl]	*PL	ID(N)	*SOC	ID(P)	*2	MAX(P)	*1	*3	*SG	MAX(N)
am[1sg]		*!		*			*		*	
art[2sg]		*!		*	*				*	
is[3sg]		*!		*				*	*	
is[sg]		*!				*			*	
are[pl]	*!					*				
☞ are[]						*				*
are[1pl]	*!			*			*			
are[2pl]	*!			*	*					
are[3pl]	*!							*		

both grammars always generate the categorical output distribution of Standard English. The temporary perturbation of ranking values during stochastic evaluation produces variant tableaux, including those shown in Figure 10.

Figure 11: Constraint Rankings on a Continuous Scale



Where do the numerical ranking values of the constraints come from? The earlier grammar was learned by the Gradual Learning Algorithm (Boersma 1997, 1999a; Boersma and Hayes 2001), starting from an initial state grammar in which all constraints have the same ranking values (arbitrarily set to be 100.0). The learning data consisted of 20,000 input-output pairs having the categorical distribution of Standard English, as shown in the first three columns of Figure 1. That is, in the learning data 100% of the outputs for [1SG] were *am*; 100% of the outputs for [2SG] were *are*, etc. The later grammar was learned by exposure to larger quantities of categorical data, given the earlier grammar as the initial grammar.[13]

For each learning datum (a given input-output pair), the GLA compares the output of its own grammar for the same input; if its own output differs from the

[13]In the Praat software used to learn the grammars (Boersma and Weenink 2000), the following parameters were set. For the earlier grammar the initial plasticity = 1, the plasticity decrement = 0.1, the number of plasticities = 2, the replications per plasticity = 10,000, the relative plasticity noise = 0.1, and the number of chews = 1. The later grammar differed by having 4 plasticities and 100,000 replications per plasticity.

given output, it adjusts its grammar by moving all the constraints that disfavor its own output upward on the continuous ranking scale by a small increment, and moving all constraints that disfavor the given output downward along the scale by a small decrement. This causes the constraints that disfavor its own erroneous output to be more strictly followed, since violations of those constraints will become more damaging the more highly ranked they are; likewise, constraints that disfavor the given output become less damaging.

Under stochastic evaluation (producing the outputs for each input in 100,000 trials), the same grammars generate the output distributions shown in Table 1. (We list in this table only outputs whose frequency is over 1/5 of 1%.) As we see in the table, stochastic evaluation produces variable outputs for the same input (e.g. "(I) am/are", "(you) are/art", "(she) is/are"), even when the learning data are categorical.

Table 1: Output Distributions of Grammars of Figure 11

Output Distributions (Outputs > 1/5 of 1%)				
input	output	% (ordinal)	% (stochastic)	
			Earlier	Later
[1SG]	am[1SG]	100	99.065	99.998
	is[SG]	0	0.234	
	are[ ]	0	0.701	
[2SG]	art[2SG]	0	0.389	
	are[ ]	100	99.512	99.996
[3SG]	is[3SG]	100	99.413	99.997
	are[ ]	0	0.503	
[1PL]	are[ ]	100	99.424	99.997
	are[1PL]	0	0.487	
[2PL]	are[ ]	100	99.782	100.000
[3PL]	are[ ]	100	99.531	99.998
	are[3PL]	0	0.444	

The amount of variation in the output depends on the standard deviation in ranking variance and the ranking distance between constraints. The earlier and later standard grammars have the same crucial ordinal constraint rankings and the same standard deviation (arbitrarily set to be 2.0, given the arbitrary units of the scale), but they spread the constraints out differently on the ranking scale. Stochastic evaluation produces less variation in the later grammar because the crucial constraints are spread farther apart. This comparison shows that stochastic evaluation can produce more or less nearly categorical output distributions depending on the ranking distance between constraints on the continuous ranking scale. The variation approaches zero as the ranking distance between the crucial constraints increases. When the ranking difference between two crucial constraints is 5 standard deviations (= 10 here), the outputs are near-categorical (0.02% reversals); when the distance is 9 standard deviations, reversal reduces to 1 in 10 billion—less than once in a lifetime (Boersma and Hayes 2001: 50).

Given the StOT framework, variation is latent in every grammar. As noted above, the later grammar was learned by exposure to larger quantities of categorical data, given the earlier grammar as the initial grammar. If the learning data had been variable instead of categorical, variable outputs could have been maintained and enhanced, and after sufficient learning the output distribution could have converged on the frequencies of variation in the learning data (Boersma and Hayes 2001).

3.3. Restrictiveness of the Constraint Set

Whether learning actually does converge on a grammar for a given set of (categorical or variable) data depends on the theory of linguistic structure embodied in the constraint set. Any given constraint set will fail to generate many hypothetical, logically possible output distributions.

Although the present constraint set is preliminary and incomplete, the space of typological possibilities it generates is orders of magnitude smaller than the space of formal patterns allowed by pure combinatorics. To see this, consider the set of abstract patterns shown in Figure 12.[14]

Figure 12: Minimal Contrast Patterns of Bare Overt Forms

a	d	a	d	a	d	a	d	a	d
b	d	a	d	b	d	b	d	a	d
c	d	b	d	b	d	a	d	a	d
a	a	a	a	a	a	a	a	a	a
b	a	a	a	b	a	b	a	a	a
c	a	b	a	b	a	a	a	a	a
a	b	a	b	a	b	a	b		
b	b	a	b	b	b	b	b		
c	b	b	b	b	b	a	b		
a	c								
b	c								
c	c								

What is shown are abstract 6-celled patterns of overt contrast consisting of up to three contrasts in the first column and none in the second column. We abstract

[14]We thank Dylan Herrick (personal communication) for suggesting Figure 12 as the means of providing a combinatorial analysis of our data.

away from the specific morphs present in each inventory by using arbitrary symbols a, b, c, d. The symbols represent only the contrasts in each paradigm; for example, a pattern consisting of a column of all a's and a column of all b's is considered equivalent to a pattern consisting of an a-column and a c-column, and the two are not counted separately. Figure 12 shows that there are 15 such minimal patterns of overtly contrasting bare forms.

The elements of our inventories, however, are not bare overt forms (phonological or orthographic entities), but *morphosyntactic forms*. Morphosyntactic forms are *tuples* of properties including both phonological and syntactic features. Simplifying, we represent a morphosyntactic form by an orthographic string plus a feature matrix for number and person: *is*[3SG], *are*[ ], etc.; compare Figure 5. Continuing to abstract away from spellings/pronunciations, we see that each pattern in Figure 12 could have a variety of different feature matrices associated with the overt forms: for example, a = [ ], b = [3SG]; a = [PL], b = [3SG]; a = [3PL], b = [3SG]; a = [ ], b = [SG], etc. The possibilities expand exponentially with the number of contrasting elements: if there are n possible matrices for one element in a pattern, then for two contrasting elements in a pattern, there are $n \cdot n (= n^2)$ possible choices of matrices. In general, for each pattern there are n^m morphosyntactic realizations, where n is the number of different feature matrices available and m is the number of distinct elements in the pattern.

We could reduce the total number of possible patterns to 15 if we (i) assumed perfect faithfulness between the input and the candidates' morphosyntactic features and (ii) simply stipulated feature homonymies for forms (so that, for example, the pattern consisting of all a's could be analyzed as in (7)):

(7)		
	a[1SG]	a[1PL]
	a[2SG]	a[2PL]
	a[3SG]	a[3PL]

But this approach would merely describe the patterns in terms of arbitrary, stipulated homonymies. The assumptions of stipulated homonymy and perfect faithfulness would deprive us of a means for explaining the extension and retraction of forms by morphosyntactic feature neutralization and generalization. Inflectional changes may arise from morphosyntactic feature simplification independently of 'accidental' form simplification caused by phonological erosion (Kusters 2000, 2001).

Therefore, we assume that our paradigms are not based on stipulated homonymies and we allow candidate feature structures to be unfaithful to the input (that is, to diverge in content from the paradigm cell they occupy). This means that forms can spread through morphosyntactic feature neutralization or generalize through remorphologization of syntactic features, as discussed below.

Suppose that we confine ourselves to just the nine feature matrices most descriptive of our data:[ ], [SG], [PL], [1SG], [2SG], [3SG], [1PL], [2PL], [3PL]. (These are the six full person-number matrices and all generalizations of these matrices in which person implies number.) Then under our assumptions of possible unfaithfulness and no stipulated homonymies, we still have the following number of logically possible morphosyntactic realizations of the agreement patterns in Figure 12 (ignoring differences in spelling or pronunciation of the forms):

$$1 \cdot (9^4) + 6 \cdot (9^3) + 7 \cdot (9^2) + 1 \cdot (9^1) = 11,511$$

As in the above observation that each pattern has n^m realizations, 9 is the number of available feature matrices n and the exponents (4,3,2, and 1) are the numbers of distinct elements m in the patterns. The coefficients (1,6,7, and 1) are the numbers of different patterns having m distinct elements; in Figure 12 these patterns are boxed together.

In contrast, our constraint set has a factorial typology of 257 different (categorical) output patterns with these 9 candidates, ignoring fixed subhierarchies.[15] That typology includes many inventories with person contrasts in plurals; if these are eliminated, the number of output patterns is reduced to 51.[16]

3.4. Relation to the Syntax of Subject-Verb Agreement

The possibility of unfaithful output expressions raises the question of how our morphosyntactic constraint set is interpreted within the larger grammar of English verb agreement.

At the level of syntax, the model is the same as in Figure 5: the INPUT is a language-independent characterization of the space of possible lexical and grammatical contrasts at the syntactic level, and the candidate set consists of language-specific syntactic structures with varying analyses of that content. As shown in Figure 13, the content is represented again by a complex feature structure (f-structure) and the candidate analyses by pairs of corresponding c(ategorial)-structures and f-structures. The INPUT feature structure contains only semantically relevant features; thus GF (for 'grammatical function') denotes any argument of the predicator (BE, in Figure 13).[17]

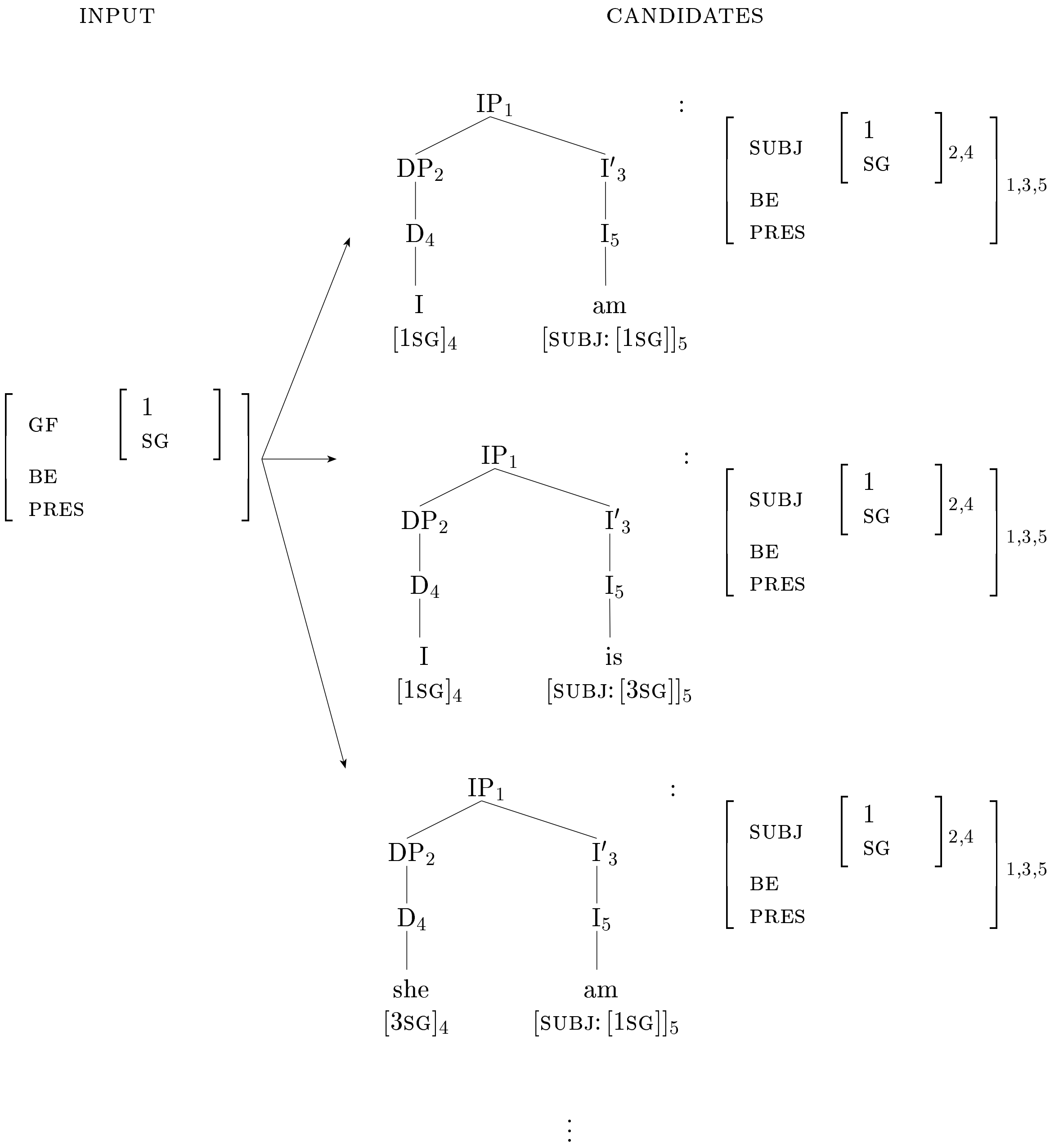
We assume that the INPUT is an underspecified f-structure which semantically subsumes the candidate f-structures, an assumption justified by considerations of decidability and learnability (Kuhn 2001). GEN provides additional purely grammatical features as well as particular argument realizations (SUBJ, for example) to the candidate analyses, which thus contain the INPUT. The terminal string of the c-structure

[15] This figure was determined by using the OTSoft software of Hayes, Tesar, and Zuraw (2000).

[16] Note that our constraint set does not by itself account for the elimination of all person contrasts in the plural. All English varieties share a great number of rankings, as well as additional constraints, which further restrict the number of theoretically possible inventories.

[17] As observed in Bresnan (in press a), an underspecified f-structure is a formal representation of the idea that the OT INPUT for syntax is an argument structure with annotations of additional semantically relevant information (Legendre, Raymond, and Smolensky 1993, Grimshaw 1997a). One advantage of this formalization is the availability of generation and parsing algorithms, recursive enumeration of the candidate set, a formal constraint language, and other useful computational and mathematical properties (Kuhn 1999b, 2000, 2001, in press). Another advantage is the typological expressiveness of the theory of representations (Bresnan 2001b).

Figure 13: OT Grammar of English Subject-Verb Agreement



GEN: INPUT→ CANDIDATES
EVAL: CANDIDATES → OUTPUT

consists of fully inflected words which represent morpholexical choices to be optimized against the candidate f-structure. The lexical choices of the sentence are optimized in parallel, so that in Figure 13 both the subject pronoun and the verb must be optimized against the given features [1SG] belonging to the SUBJ argument of the candidates. Lexical choices may be unfaithful to the candidate f-structure/input, as already illustrated in Figure 5.[18]

Since the candidate feature structures are all semantically subsumed by the input in this model, the lexical optimizations can be carried out against the candidate f-structure, which in general contains the input together with purely grammatical features provided by GEN. More precisely, then, the faithfulness constraints defined previously will relate the morpholexical f-structures of the c-structure terminals to the global feature structures of the candidates. Again, different lexical optimizations (for example, those for the subject pronoun and the verb) may proceed in parallel.

Now different expressions in the lexical string may conflict in their featural specifications, as when a first person subject pronoun cooccurs with a third person verb, as found in Yorkshire, Figure 1 (1) and Northumberland Figure 1 (4). In general, faithfulness to the referentially classificatory feature of person is much stricter for pronominal expressions than for verbal expressions. This point is illustrated by the fact that in Figure 13, the first two candidates *I am* and *I is* are both possible expressions of the input with its first person singular argument, while the third candidate *She am* is always suboptimal. (Note that *She am* is an optimal expression of a *third* person subject in some English varieties, as noted above; we suggest that it is suboptimal only as an expression of a *first* person subject.) This generalization can be captured by the subhierarchy (8):

[18]It may be useful for those unfamiliar with feature-logic basic theories of syntactic representation to view this as a feature checking system without movement—output oriented rather than derivational. The basic workings of the system of feature-structure comparison are as follows. The numerical subscripts coindexing the tree nodes and feature structures show the correspondence relations between the two parallel structures, which follow from general principles of tree-to-feature-structure correspondence (Bresnan 2001b, Kuhn 1999a). For example, the feature structures associated with the I nodes in these particular trees are indexed by 5, which is identified with the index of I′ (=3) and IP (=1) by a principle that identifies the f-structures of heads with those of their mothers. Similarly, the feature structures of the D nodes are indexed by 4, which is identified with the index of DP (=2) by the same head principle. The DP and IP f-structures are related by the specifier principle, which says here that f-structure 5's SUBJ has f-structure 2 as its value. (Other principles apply to the exocentric and nonconfigurational constructions found in many languages; see Bresnan 2000b, Nordlinger 1998.)

In faithfulness evaluations, the lexical feature structure of a terminal node is compared with the f-structure corresponding to (coindexed with) its preterminal node in the c-structure. By the syntactic correspondences in Figure 13 just discussed, this comparison will hold for the f-structures of the phrasal projections of these terminals (IP in the case of *am*, *is*, and DP in the case of *I, she*). By the uniqueness principle, which states that every f-structure attribute must have a unique value, the verb's inner agreement feature structure [1SG] in $[\text{SUBJ}\ [1\text{SG}]]_5$ can be inferred to correspond to the subsidiary f-structure 4 (=2) in the sentential feature structure, which also corresponds to the lexical feature structure of the subject pronoun. For more details of the LFG representational basis adopted here, see Bresnan (2001b) and references.

(8) Constraint subhierarchy for expressions of person:

$\text{FAITH}_{\text{PRON}}$ (PERS) $\gg$ $\text{FAITH}_{\text{VERB}}$ (PERS)

The subscripted faithfulness constraints in (8) are again positional faithfulness constraints, indexed respectively to the morphosyntactic domains of pronominal and verbal expressions. $\text{FAITH}_{\text{VERB}}$ (PERS) stands for the faithfulness constraints we have already discussed above: MAX and IDENT, which are indexed to specific verb classes. $\text{FAITH}_{\text{PRON}}$ (PERS) stands for corresponding faithfulness constraints for pronominals, which also have varying form classes (including some pronominal inflections bound to verbs, Bresnan in press c). The verbal and pronominal positional faithfulness constraints are separately rankable, but the subhierarchy (8) ensures that the subject pronoun cannot be less faithful to the input person of the subject argument than the verb is.[19]

This theory is the basis for our analysis of the inter-speaker and intra-speaker variation in *be* inventories extracted from the *SED*.

4. Analysis of the Inventories

Here we apply our stochastic OT theory to issues in the analysis of the categorical and variable data paradigms extracted from the *SED*.

4.1 Feature Spreading or Deletion?

As we will now show, both feature spreading and feature deletion are possible mechanisms in the extension of forms. The former preserves the morpholexical inventory, but leads to purely formal agreement features. The latter preserves the semantic content of features, but changes (remorphologizes) the inventory.

Yorkshire The "Yorkshire" inventory in Figure 1 (1), which stands for a paradigm found in Yorkshire/Lancashire/Westmoreland/Northumberland, expresses all persons in the singular with *is*, *are* being used in the plural. If we assume that *is*, *are* in Yorkshire have the same content that they do in Standard English, this means that *is*[3SG] is the optimal candidate for [1SG] and [2SG] as well as [3SG] inputs, and *are*[ ] for all plural inputs. This Yorkshire inventory is derived by the constraint ranking in Figure 14.

This Yorkshire grammar differs crucially from that for Standard English in Figure 10 in having the rankings *1 $\gg$ IDENT(PERS) and MAX(PERS) $\gg$ IDENT(PERS). These rankings mean that first person will be highly marked, like second person in Standard English, but the avoidance of the marked persons will take a different route. Where the Standard English grammar favors deleting the second person feature

[19] It is notable that unlike person, number is a category in which pronominal expressions may be less faithful than verbal expressions. For example, in Golin, a Papuan language of New Guinea, both bound and free pronouns are undifferentiated for number contrasts but there is a verbal suffix specialized for first person singular subjects (Foley 1986: 70).

Figure 14: Tableaux of a Grammar of Yorkshire English

[1sg]	ID(N)	*PL	*2	*1	MAX(P)	MAX(N)	*SOC	*SG	ID(P)	*3
am[1sg]				*!				*		
art[2sg]			*!					*	*	
☞ is[3sg]								*	*	*
is[sg]					*!			*		
are[pl]	*!	*			*					
are[]					*!	*				
are[1pl]	*!	*		*						
are[2pl]	*!	*	*						*	
are[3pl]	*!	*							*	*

[2sg]	ID(N)	*PL	*2	*1	MAX(P)	MAX(N)	*SOC	*SG	ID(P)	*3
am[1sg]				*!			*	*	*	
art[2sg]			*!				*	*		
☞ is[3sg]							*	*	*	*
is[sg]					*!		*	*		
are[pl]	*!	*			*					
are[]					*!	*				
are[1pl]	*!	*		*					*	
are[2pl]	*!	*	*							
are[3pl]	*!	*							*	*

[3pl]	ID(N)	*PL	*2	*1	MAX(P)	MAX(N)	*SOC	*SG	ID(P)	*3
am[1sg]	*!			*				*	*	
art[2sg]	*!		*					*	*	
is[3sg]	*!							*	*	*
is[sg]	*!				*			*		
are[pl]		*!			*					
☞ are[]					*	*				
are[1pl]		*!		*					*	
are[2pl]		*!	*						*	
are[3pl]		*!								*

over changing its value (enforced by the ranking IDENT(PERS) ≫ MAX(PERS)), this Yorkshire grammar favors replacement over deletion (recall n. 9).

The reason that *is* does not spread to the plural in this Yorkshire grammar is that IDENT(NUM) dominates MAX(NUM) in the constraint ranking, meaning that it is worse to rewrite feature values (violating IDENT) than to delete them (violating MAX). Complete spreading of *is*[3SG] throughout both singular and plural would result from the reverse constraint ranking of these two constraints in Figure 14.

As we see from this example, IDENT violations can cause the 'rewriting' of input feature values in the output. The use of *is*[3SG] for all singular persons in the Yorkshire grammar implies that the person value of the output is not part of the *content* of the sentences containing it, represented by the input. Such unfaithful features of the candidate are purely formal features, in the sense that they are unrelated to content. In general, morphosyntactic IDENT violations will produce such divergences between form and content. An example from Romance clitic inventories where number and gender features "float" onto adjacent clitics in certain circumstances (Bonet 1995) is analyzed in terms of IDENT violations by Grimshaw (1997b: 193–4; in press). When the divergence between the form and content of the candidate is contextually restricted, as in the Romance example, the output alternates between a faithful form and an unfaithful form that replaces it in limited circumstances. The contentful features of the input are thus only contextually neutralized, and are still transparent in most output forms. But the Yorkshire grammar of Figure 14 gives us absolute (context-free) neutralization of person in the output, so that the candidate's person feature could be opaque in every context of its use.[20] In this situation "remorphologization" of the system may occur, replacing the candidate's unfaithful features with a more faithful, and therefore meaningful, analysis.

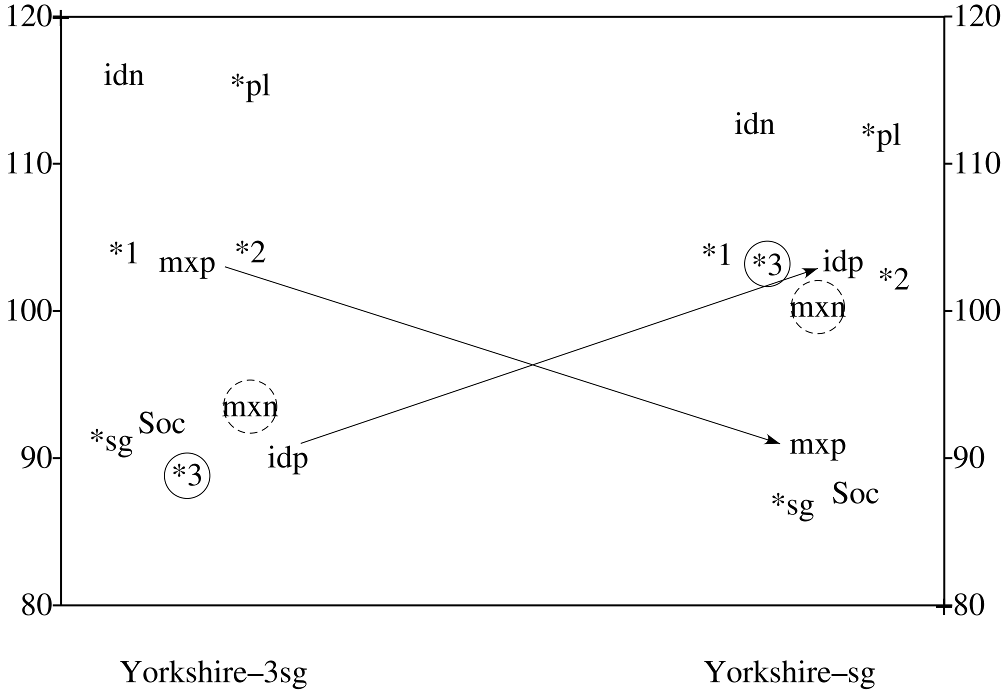
In the present theory, constraint reranking can cause the 'remorphologization' of the inventory. This process is illustrated in Figure 15. The grammar on the left is that of Figure 14, for which *is*[3SG] is the optimal output for all singular persons. The grammar on the right shows the rerankings that make *is*[SG] the optimal output. The key rerankings are the reversal of MAX(PERS) and IDENT(PERS), which favors deleting person over changing person values, and the elevation of *3 above MAX(PERS), which favors deleting third person as well as first and second.

Tableaux of the resulting grammar are shown in Figure 16. Observe that the remorphologized Yorkshire grammar differs from the Standard English grammar not only in constraint ranking, but in the content of the *be* inventory: where *is* in Standard English has [3SG] content, the remorphologized Yorkshire *is* is 'impersonal', having only [SG] content.

Our data are insufficient for us to infer the exact historical path taken by the English variety spoken by the "Yorkshire" respondents. The 'remorphologized' York-

[20]Because our data set, like our constraint set, is small and incomplete, we cannot of course be certain that there are not relevant alternations elsewhere in the grammar. Indeed, the "Northern Rule" affecting verb agreement when a subject pronoun is adjacent would be relevant in some Yorkshire inventories, as Andrew Garrett (p.c.) reminds us; see Börjars and Chapman 1998. All of our *SED* inventory verbs come from sentences with pronoun subjects.

Figure 15: Remorphologization: is[3SG] → is[SG]



shire grammar shown in Figure 15 as "Yorkshire-sg" was learned directly from the *is*[SG] distribution, without passing through any stage that conserves the featural content of Standard *is*. However, remorphologization could equally well come from taking the "Yorkshire-3sg" grammar as the starting point. (Likewise, "Yorkshire-sg" could be the starting point for learning "Yorkshire-3sg".)

The point of interest here is that gradual changes on the continuous ranking scale can give rise to categorical changes in content—without any derivational operations or procedures.

By the same line of reasoning, the analysis of *are* as a general form lacking PERS and NUM features may be the result of historical remorphologization of an earlier more specific plural form. In the Yorkshire and Derbyshire/Cornwall inventories, as we have seen in Figure 1 (1) and (7), *are* is restricted to the plural. But elsewhere in our data *are* generalizes into the singular column of the paradigm, expressing the second person or both second and first persons. An example is the Northumberland inventory in Figure 1 (4).

Northumberland The Northumberland inventory—*I is, you are, she is, we/you/they are*—differs from Yorkshire in Figure 1 (1) in having *are* for [2SG] input. Reranking the 'social deixis' constraint SOC above MAX(NUM) in either Yorkshire grammar

Figure 16: Tableaux of ‘Remorphologized’ Yorkshire Grammar

[1sg]	ID(N)	*PL	*3	*1	ID(P)	*2	MAX(N)	MAX(P)	*SOC	*SG
am[1sg]				*!						*
art[2sg]					*!	*				*
is[3sg]			*!		*					*
☞ is[sg]								*		*
are[pl]	*!	*						*		
are[]							*!	*		
are[1pl]	*!	*		*						
are[2pl]	*!	*			*	*				
are[3pl]	*!	*	*		*					

[2sg]	ID(N)	*PL	*3	*1	ID(P)	*2	MAX(N)	MAX(P)	*SOC	*SG
am[1sg]				*!	*				*	*
art[2sg]						*!			*	*
is[3sg]			*!		*				*	*
☞ is[sg]								*	*	*
are[pl]	*!	*						*		
are[]							*!	*		
are[1pl]	*!	*		*	*					
are[2pl]	*!	*				*				
are[3pl]	*!	*	*		*					

[3pl]	ID(N)	*PL	*3	*1	ID(P)	*2	MAX(N)	MAX(P)	*SOC	*SG
am[1sg]	*!			*	*					*
art[2sg]	*!				*	*				*
is[3sg]	*!		*		*					*
is[sg]	*!							*		*
are[pl]		*!						*		
☞ are[]							*	*		
are[1pl]		*!		*	*					
are[2pl]		*!			*	*				
are[3pl]		*!	*							

causes *are* to replace *is* as the expression of second person singular input. All other outputs are unaffected, because Soc only marks singular expressions of second person input. The tableau in Figure 17 illustrates the outcome. It differs crucially from the [2sg] tableau in Figure 16 only in the ranking of Soc.

Figure 17: Tableau of Northumberland Grammar: [2sg] → *are*[ ]

[2sg]	Id(n)	*pl	*Soc	*3	*1	Id(p)	*2	Max(n)	Max(p)	*sg
am[1sg]			*!		*	*				*
art[2sg]			*!				*			*
is[3sg]			*!	*		*				*
is[sg]			*!						*	*
are[pl]	*!	*							*	
☞ are[]								*	*	
are[1pl]	*!	*			*	*				
are[2pl]	*!	*					*			
are[3pl]	*!	*		*		*				

An alternative ranking which yields *are*[pl] for plurals and second person rather than *are*[ ] is also consistent with the data given so far. However, *are*[pl] is inconsistent with further data, to be discussed below.

Kent Another case where an extension of plural form into the singular column implies remorphologization is the Kent inventory in Figure 1 (2)—*I/you are, she is, we/you/they are*. Here *are* has generalized to first and second person singular. This distribution cannot be captured within our present theory, if *are* is analyzed as in the inventory on the right in (9):

(9) Analyses of Kent, Figure 1 (2) in the present theory:

Possible:

	sg	pl
1	*are*[ ]	*are*[ ]
2	*are*[ ]	*are*[ ]
3	*is*[3sg]	*are*[ ]

Impossible:

	sg	pl
1	*are*[pl]	*are*[pl]
2	*are*[pl]	*are*[pl]
3	*is*[3sg]	*are*[pl]

For *are*[pl] to win over candidates faithful to person, Max(pers) must be low-ranked; for it to win over the general candidate *are*[ ], Max(num) must be ranked higher than *pl. The remaining competitor is *is*[sg]. But only three constraints distinguish *are*[pl] from *is*[sg]: Ident(num), *pl, and *sg. The six possible rankings of these constraints give three different outcomes for these two competitors. If *pl ≫ Ident, *sg, then *is*[sg] uniformly wins on all input; if *sg ≫ Ident, *pl then *are*[PL] uniformly wins on all input; and if IDENT ≫ *PL, *SG, then the output splits on singular and plural input, as shown in (10). There is no ranking of these constraints which yields the Kent inventory.

(10) Competition between *are*[PL] and *is*[SG]:

/[1sg]/	IDENT(N)	*PL	*SG
are[pl]	*!	*	
☞ is[sg]			*

/[1pl]/	IDENT(N)	*PL	*SG
☞ are[pl]		*	
is[sg]	*!		*

In contrast to *are*[PL], the general form *are*[ ] can generalize part-way into the singular column, as shown by the grammar in Figure 18.

We see from the above analyses that the generalization or spread of a form in the *be* paradigm can proceed in the present theory by (the OT equivalents of) either feature deletion or feature change, and that the generalization of *are* across both number and persons in some dialects requires the deletion analysis, under which it lacks both PERS and NUM values.

4.2. Relations between Verb Classes

Further evidence bearing on the choice of analyses can be found by comparing main verb agreement patterns with those of *be*. The inventories extracted for the present study do not include main verb agreement, but we can illustrate the point with evidence from Standard English.

Standard It has been argued that *are* is a general form (having the agreement matrix [ ]) on the basis of its generalization to first person singular in negative interrogative inverted positions (Bresnan 2000, 2001a, in press a):

(11) Aren't I your friend?

By the same reasoning just given for the Kent grammar, (11) supports *are*[ ] in Standard English. The restriction in Stndard English to inverted position (*Aren't I your friend, *I aren't your friend*) is explained by the syntactic theory given in Bresnan (in press a).

Further evidence for *are*[ ] as the general form in Standard English comes from the agreement pattern of main verbs. English main verbs make fewer person contrasts in the singular than *be*; in fact, the main verb paradigm matches the Kent paradigm, Figure 1 (2) in abstract agreement contrasts:

Figure 18: Tableaux of Kent Grammar

[1sg]	ID(P)	*1	*PL	*SOC	*2	ID(N)	MAX(P)	*SG	MAX(N)	*3
am[1sg]		*!						*		
art[2sg]	*!				*			*		
is[3sg]	*!							*		*
is[sg]							*	*!		
are[pl]			*!			*	*			
☞ are[]							*		*	
are[1pl]		*!	*			*				
are[2pl]	*!		*		*	*				
are[3pl]	*!		*			*				*

[2sg]	ID(P)	*1	*PL	*SOC	*2	ID(N)	MAX(P)	*SG	MAX(N)	*3
am[1sg]	*!	*		*				*		
art[2sg]				*!	*			*		
is[3sg]	*!			*				*		*
is[sg]				*!			*	*		
are[pl]			*!			*	*			
☞ are[]							*		*	
are[1pl]	*!	*	*			*				
are[2pl]			*!		*	*				
are[3pl]	*!		*			*				*

[3sg]	ID(P)	*1	*PL	*SOC	*2	ID(N)	MAX(P)	*SG	MAX(N)	*3
am[1sg]	*!	*						*		
art[2sg]	*!				*			*		
☞ is[3sg]								*		*
is[sg]							*!	*		
are[pl]			*!			*	*			
are[]							*!		*	
are[1pl]	*!	*	*			*				
are[2pl]	*!		*		*	*				
are[3pl]			*!			*				*

(12) Kent *be* paradigm:

	SG	PL
1	*are*	*are*
2	*are*	*are*
3	*is*	*are*

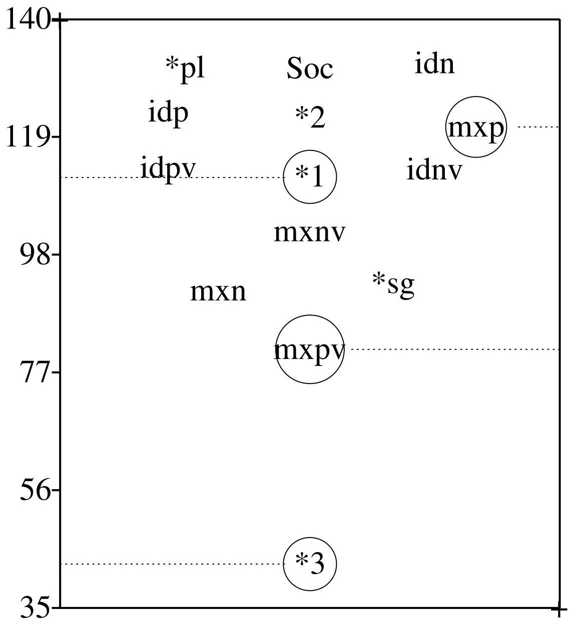
Standard main verb paradigm:

	SG	PL
1	*run*	*run*
2	*run*	*run*
3	*runs*	*run*

Hence, the same ranking arguments support the subject agreement matrix [ ] for both *run* and *are*. (See the discussion of (9) and (10)).)

Now recall (with reference to the discussion of examples (2)–(4)) that the faithfulness constraints for person and number are indexed to the distinct inflectional classes of verbs, in accordance with the theory of positional faithfulness. It follows that the faithfulness constraints for *be* and main verbs must be consistent with the same ranking of markedness constraints. There is in fact a consistent ranking that produces *run*[ ] and *are*[ ] as the non-third-singular agreement forms for *be* and main verbs. This ranking is shown in Figure 19.

Figure 19: Positional Faithfulness to Person
for *be* (mxp) and V (mxpv) in Standard English



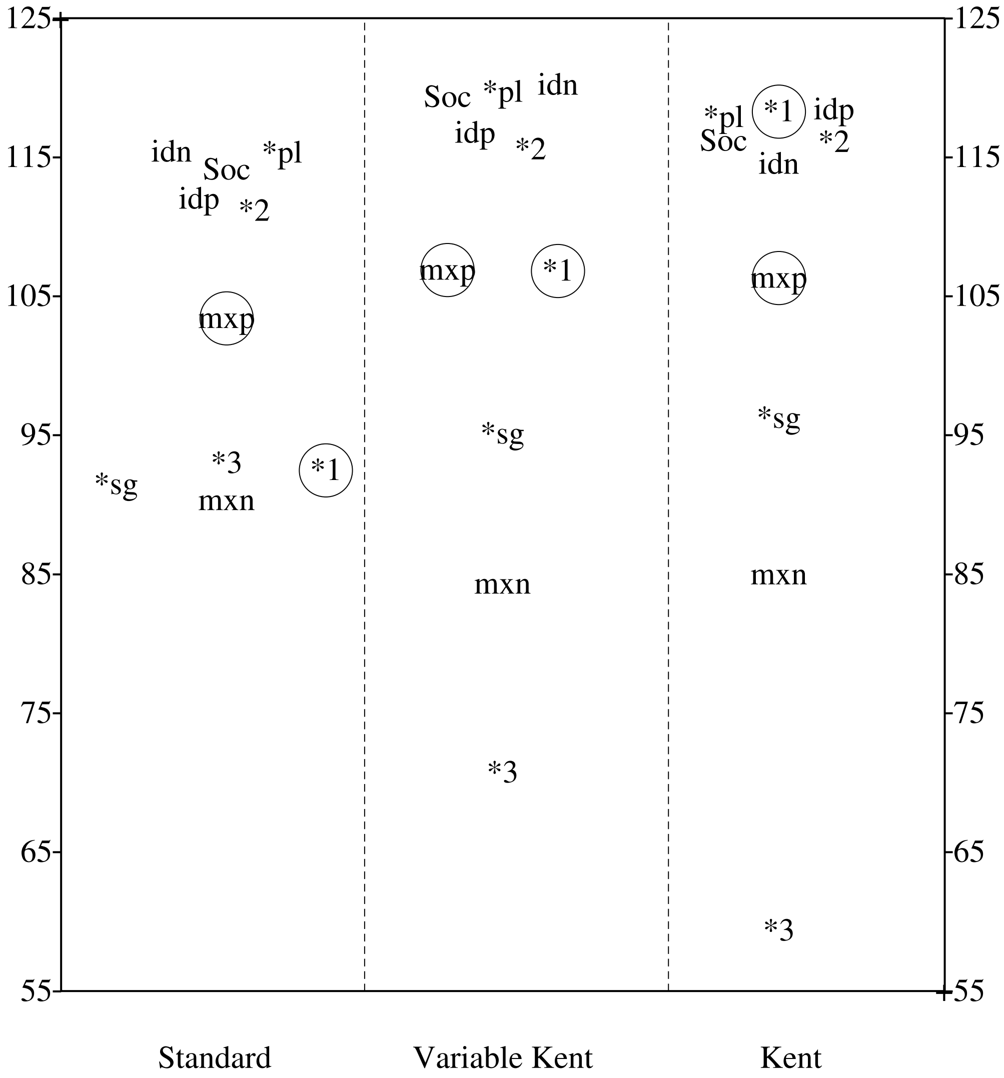
4.3. Relations between Variable and Categorical Outputs

An interesting property of the stochastic OT framework is that similarity of grammatical structure between two varieties maps into distances in the constraint space.

This property allows us to depict graphically the similarities of a variable variety to several categorical varieties.

Variable Kent The variable Kent/Bedfordshire inventory Figure 2 (10) combines the Kent and the Standard patterns. The relations of the three grammars are shown in Figure 20, where the crucial faithfulness and markedness constraints that determine the choice of *are* or *am* for [1SG] input are circled.

Figure 20: Variation as a Function of Ranking Distane: Variable Kent



In Standard English the distinctive [1SG] form *am* is admitted into the inventory by the very low ranking of *1 below MAX(PERS), as we saw earlier. In contrast, the categorical Kent variety, where *am* is replaced by *are*, has just the reverse ranking, with *1 well above MAX(PERS). Observe that in the variable Kent grammar of Figure 20 *1 is midway in ranking position between its position in the categorical grammars of Standard and Kent English.

The relative frequency of the alternative outputs *am* and *are* in the variable Kent grammar is a function of the ranking distance between these two constraints. Because we lack frequency information for the *SED* inventories, we simply assume that each is used 50% of the time. The variable Kent/Bedfordshire grammar illustrated has approximately this distribution.

Derbyshire and Variable Yorkshire Derbyshire/Cornwall inventory in Figure 1 (7) has a distinctive [2SG] form *art*: *I am, you*[21] *art, she is, we/you/they are.* This inventory differs from the Standard only on second person singular input. Where the Standard grammar rates the markedness of the second person above the faithfulness to person, Derbyshire/Cornwall ranks the constraints that mark second person singular—*2, SOC—below MAX(PERS). This difference is shown in the leftmost and rightmost grammars of Figure 21.

The variable Yorkshire inventory in Figure 2 (11) differs from the Derbyshire/Cornwall inventory, Figure 1 (7), only in having both *art* and *are* for the same [2SG] input. Observe in Figure 21 that in the Derbyshire/Cornwall grammar the markedness constraint against second singular SOC (the "social deixis" reflex) is ranked very close to MAX(PERS). The close ranking distance between these two constraints produces variations in the output through constraint dominance reversal under stochastic evaluation.

Other Variable-Categorical Relations All of the other relations between variable and categorical inventories in Figures 1 and 2 can be analyzed in the same way: the variable grammars have crucial constraints at an intermediate ranking distance between those of the related categorical patterns. Figures 22–24 summarize the rankings.

4.4. Conclusions from the Analyses

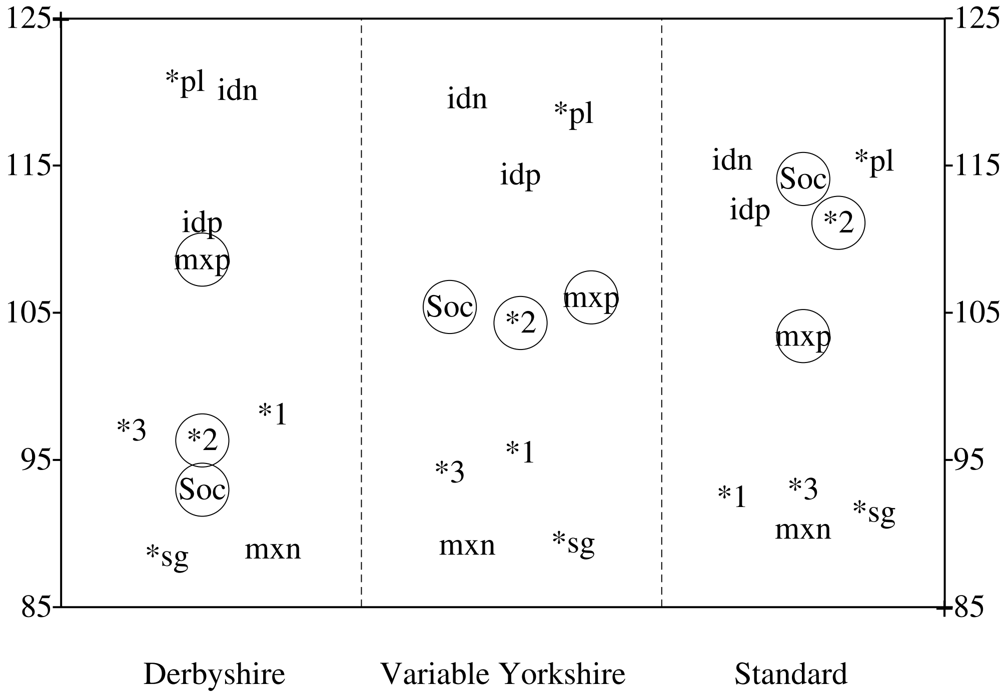
From our analyses of these varieties we draw the following conclusions.

Conclusions from the Categorical Inventories:

- both morphosyntactic spreading (via feature change, or neutralization) and

[21] The *SED* does not elicit second person singular pronouns in unemphatic affirmative sentences. Descendents of *thou, thee* occur in traditional dialects in the north and west of England (Trudgill and Chambers 1991: 7–8). In Somerset English *thee* is more frequent than *you* in stressed positions, and *you* has replaced *ee* is preverbal subject position, whether stressed or not (Ihalainen 1991: 115–116). The use of *her* as a nominative also occurs in some dialects, but all such variation in pronominal inventories lies outside the scope of the present study.

Figure 21: Variation as a Function of Ranking Distance: Variable Yorkshire



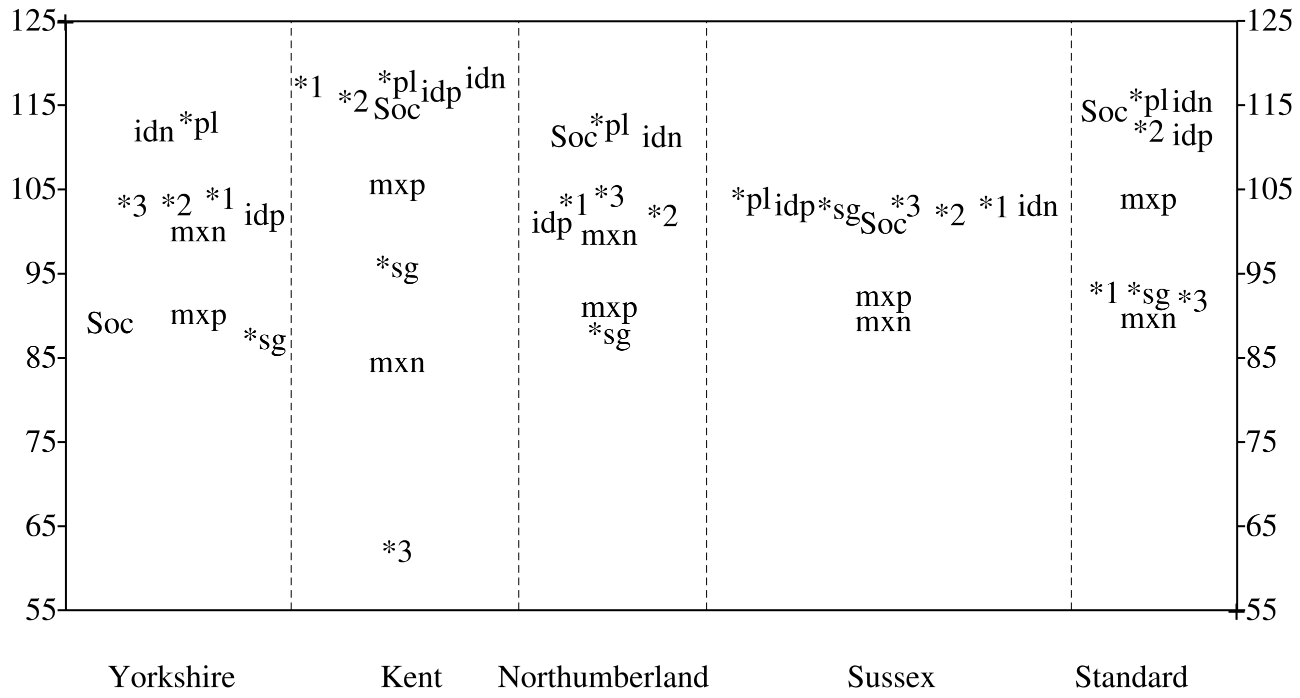
morphosyntactic generalization (via feature deletion) are possible mechanisms for the extension and retraction of agreement forms within StOT

- gradual changes on the continuous ranking scale can give rise to categorical changes in lexical content ('remorphologization')
- the analysis of *be* interacts with the analysis of other verb classes via positional faithfulness; the analysis of the general form *are*[ ] in Standard English of Bresnan (2000, 2001a, in press a) is supported

Conclusions from the Variable Inventories:

- a single OT grammar with stochastic evaluation can simultaneously generate both categorical and variable morphosyntactic outputs
- OT grammars with stochastic evaluation can explain the existence of variation between the general form and specific forms as expressions of the same input, which has been problematic for theories using blocking or preemption of defaults by more specific forms, and for ordinal OT

Figure 22: Summary of Grammars I



- similarity of grammatical structure between two varieties maps into distances in the constraint space
- there are grammatical constraints on intra-speaker variation, which are identical to constraints on inter-speaker and typological variation

4.5. Residual Inventories

We see from the preceding analyses that much of the *be* inventory variation within and across speakers can be explained as competition among candidates differing in the complexity of their featural agreement content. However, it goes without saying that not all variation can be explained in these terms, and the same is true of our data. Some of the variation must be attributed to other factors than agreement feature complexity.

The variable Durham inventory in Figure 4 (17) has *I/you/she is, we/you are, they is/are*. If we look at the similar but invariable Yorkshire grammar in Figure 16, we see that the split between *is* in the singular and *are* in the plural comes from the high priority the grammar places on preserving the value of the number feature while marking plural over singular; faithfulness to person is a lesser priority. With singular inputs, the high rank of IDENT(NUM) above MAX(PERS) selects *is* over *are*, given the low priority of faithfulness to person. With plural inputs, the high rank of

Figure 23: Summary of Grammars II

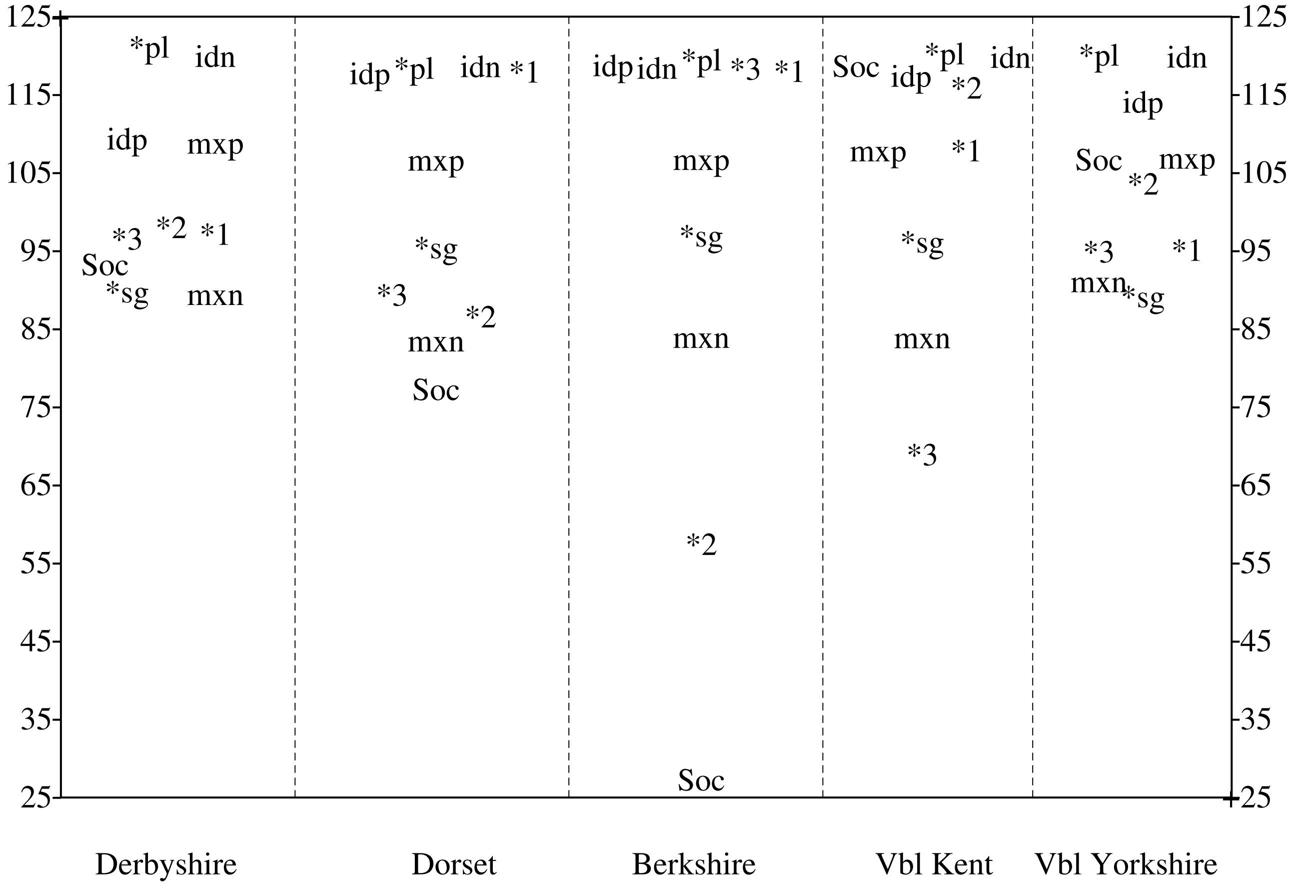
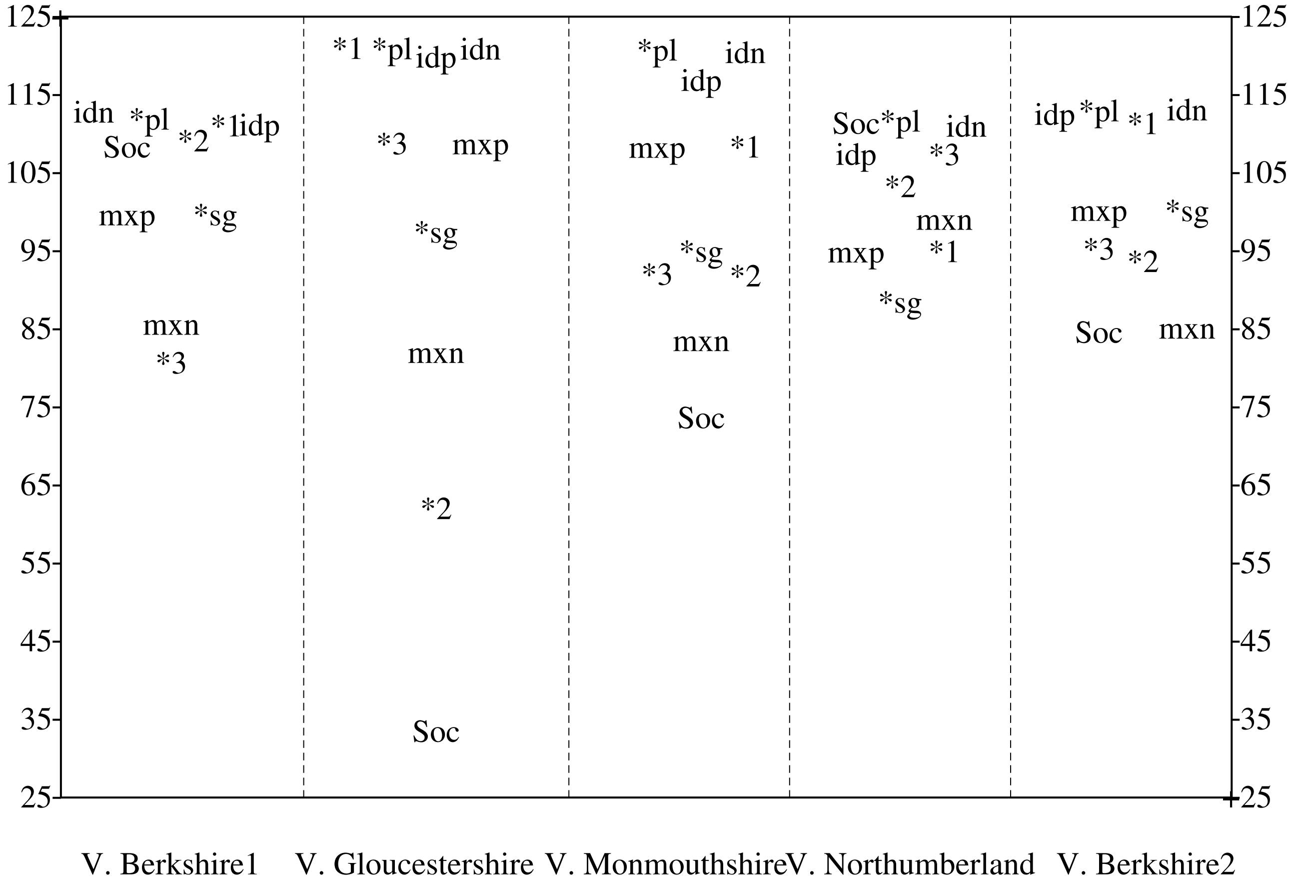


Figure 24: Summary of Grammars III



IDENT(NUM) now excludes *is*, while *PL disfavors specific plural forms, making the general form *are* optimal. None of these constraints picks out a specific person-number combination, so the variable Durham output [3PL] → *is*, *are* cannot be derived. What is derivable is the inventory in (13):

(13) Hypothetical Generalized Durham

	sg	pl
1	is	are, is
2	is	are, is
3	is	are, is

By lowering IDENT(NUM) in the Yorkshire grammar to be very close to MAX(PERS), this hypothetical "Generalized Durham" inventory is obtained, as the reader can check. The actual Durham inventory (17) cannot be derived without adding further constraints.

The *am* inventories are also problematic for our small constraint set. Under the present theory we can derive an invariant *am* inventory, but we cannot capture the particular variations shown in Figure 4 (18a,b). Here there is certainly reason to think that other constraints must be involved. Ihalainen (1991) observes that in the generalized *am* dialects in East Somerset, *am* is used as an unstressed allomorph of *be*: *'m* (Ihalainen 1991: 107–8). Of its syntactic distribution he writes:

> The enclitic *'m*, as in *we'm, you'm,* and *they'm* only occurs in unstressed position in declaratives. There are no sentences like *Am they all right?* or *Yes, they am.*
>
> The singular form *is* is frequently used with plural noun subjects, as in *These here is spares* and *All horses is gone*. However, it does not seem to occur with personal pronouns, not even with speakers who use a singular past tense form with plural subjects (*we was, you was, they was*). The only exception may be *they* in questions (*Is they all right?*).

If the generalized *am* form is an unstressed enclitic on the subject, further constraints are needed to regulate its distribution and presence in the inventory. If the choice of allomorph depends on the nominal expression type of the subject (NP or pronoun), that too requires further constraints.[22]

Apart from such phonological and morphological factors, there are also the issues of salience and prominence invoked by Ihalainen (1991) and Cheshire (1996), as well as the differential use of specific forms for social expression that has been observed in the sociolinguistic literature on *be* variation.

[22]The reduction of some of the Standard English auxiliary clitics to morphological affixes of subject pronouns has been observed in several careful morphological studies (Spencer 1991, Sadler 1998, Barron 1998, Nordlinger and Sadler 2000; cf. Bender and Sag in press), and is well-attested typologically (Nordlinger 1998, Nordlinger and Sadler 2000). See also the references in n. 20.

Thus we must leave these residual inventories unexplained, pending a fuller investigation of the data and an appropriate placement of the current partial constraint set within a broader theoretical context.

5. Implications of a Stochastic Competence Grammar

Recall that all of our variable inventories, like the categorical ones, represent the outputs of individual speakers or groups of speakers with shared responses. Given that a single StOT grammar can produce both variable and categorical outputs and explain their shared grammatical structuring, we can hypothesize that variation is part of the internalized knowledge of language—the linguistic 'competence'—of speakers of these varieties.

One objection we have encountered to the idea of a stochastic competence grammar is given in (14):

(14) How can a stochastic grammar represent an individual speaker's internalized linguistic competence? —if alternative outputs are randomly generated, the speaker cannot know what she is going to say!

This objection stems from the misconception that a stochastic process involving a probability distribution represents something intrinsically random and unknowable. We refer to it as **the Fallacy of Reified Ignorance**. In fact, the stochastic models represent gaps in our knowledge of the world, not gaps in the causal structure of the world. The speaker does of course know what she is going to say. The specific choice of variant outputs is *not* determined solely by the grammar, and stochastic evaluation provides an explicit model of this fact.

To explain the stochastic model of variation at the micro-level, Boersma provides a neural-net analogy (Boersma 2000: 483): "the loudness of the protest of a constraint is the value of an inhibitory postsynaptic potential: it depends on the synaptic strength (the ranking as specified in the grammar) as well as on some things like the accidental amount of locally available neurotransmitter." This conception is summarized in an equation, which we present in simplified form as (15):

(15) effective ranking = constraint ranking + factors not taken account of in our theory

When the factors not taken account of in our theory can be assumed to be independent of the formal grammar theory, a probability distribution is useful, and the normal (gaussian) distribution chosen by Boersma serves this purpose, as a placeholder for these other micro-level causal factors involved in variation at the neural level in his analogy. Boersma (1997, 1998, 2000) refers to this random factor as 'noise', and to the stochastic evaluation of candidates as 'noisy evaluation'.

It is well known from sociolinguistics that macro-level factors—such as the social meaning of an expression in a certain context—affect variation. While some aspects of social meaning could be grammaticalized into the contents of expressions and constraints (morphological markers of politeness levels, for example), other social aspects could be independent of the grammar fragment/partial theory in question, constituting 'noise' to the syntactician, perhaps. A third way that sociolinguistic factors could affect variation is by systematically boosting or depressing selected constraints. A model of this effect is given in (16) (adapted from Boersma and Hayes, 2001: 82–3):

(16) effective ranking = constraint ranking$_i$ + $styleSensitivity_i \cdot Style$ + noise

Here *styleSensitivity* is a constraint-specific value added to the constraint ranking: when positive, a constraint's ranking is boosted; when negative, the ranking is depressed; and when zero, the ranking is unaffected, or stylistically neutral. *Style* is a continuous variable ranging from 0 (for most casual style) to 1 (for most formal). According to this model, the rankings of various 'style sensitive' constraints may covary (directly and inversely) with the speech style. These covarying subgrammars could be viewed as representing **sociolinguistic competence**.

Grammatical constraints on variation cannot be captured within the classical generative framework except by hypothesizing separate categorical grammars for each parameter of variation (Kroch 2000: 720):

> ... variation in syntax which corresponds to opposed settings for syntactic parameters [e.g. the V2 parameter] must reflect the co-presence in a speaker or speech community of mutually incompatible grammars.

Given extensive independent variation along these parameters, this approach would require an unattractive multiplication of grammars to characterize the output of a single speaker. For example, both *you* and descendents of *thou, thee* occur in traditional dialects in the north and west of England (Trudgill and Chambers 1991: 7–8), but the forms cannot simply be consigned to different grammars, say a "dialect grammar" and a "standard grammar", because they can often be mixed within the same sentence:

(17) *You taught theeself, didn't ee?* (Ihalainen 1991: 115)

I'm not under no obligation about this, be I?, They're not ready, be 'em? (Ihailanen 1991: 109, 116)

Moreover, there is mixing of the variant pronominal forms together with the variant verbal forms, as illustrated in the following extract from taped Somerset speech (Ihalainen 1991: 115):[23]

[23]This example is used by Ihalainen to illustrate the fact that *thee* is used more frequently in stressed positions than in unstressed ones, such as *Art thee married?*.

(18) B.I. *What be you, Herb? Seventy-two?*
H.T. *Gone seventy-five.*
B.I. *Seventy-five! Thee!*
W.B. *Thee! Thee! I didn't know you were gone seventy-five.*

The competing grammar model can generate variable outputs characterized by covarying grammatical dependencies, but outputs characterized by independently varying grammatical factors lead to exponential growth of competing grammars. For example, if the grammar of a single speaker has n cases of independently varying output caused by reranking different constraints, we would have to assume 2^n competing grammars for this speaker to explain this kind of variation. If the number of independent grammatically determined variable outputs n is only 10, the number of competing invariable grammars required is over one thousand.

In contrast, a single stochastic OT grammar could easily generate this output distribution (cf. Guy 1997). StOT grammars can also generate dependent covarying syntactic outputs as we saw in our discussion of Boersma and Hayes' (2001) model of stylistic variation (16). Even the hypothesized constant rate of historical change of covarying outputs, which has been cited in support of the competing grammars approach, can also be captured within stochastic Optimality Theory (Clark 2001).

For many, linguistic competence would exclude variation by definition; for them there is a fixed conception of 'competence' that statically pre-defines the subject matter of linguistic theory. Yet as semantic theories have advanced, many phenomena that at one time were thought to lie in the realm of pragmatics outside of formal semantics, have been incorporated into formal semantic theories. Likewise, as syntactic and morphosyntactic theories advance, we should expect a corresponding enlargement of their empirical domains.

Stochastic Optimality Theory in fact predicts that the variable outputs of individual speakers' grammars should be constrained by the same kinds of categorical generalizations that are found across grammars. Typological variation across languages is explained within OT by means of constraint reranking. The stochastic evaluation of candidates in individual grammars reranks these same constraints by temporarily perturbing their ranking values along the continuous scale. As a result, individual variation samples the typological space of possible grammars.

Appendix. Limitations of the Present Study

The *SED* has certain limitations for our purposes. Some questionnaires are not completed, and the *SED* fieldworkers completed some with more than one informant. We discarded a few inconsistent-looking inventories in which the set of forms in response to Item IX.7.7 did not intersect the set of forms in response to Item VIII.9.5. Note also that there is no questionnaire item at all eliciting second person plural '[2PL]' forms, and inverted first person plural [1PL] forms in yes/no questions are also absent. We assume the [2PL] gap in responses is filled by the general form used for other plurals.

This form is attested in some studies of particular dialects. Further, questionnaire responses, like other data collected through elicitation, may inaccurately reflect the use of these forms in actual speech (Ihalainen 1991: 110; Schilling-Estes and Wolfram 1994: 297). For all these reasons, the extent of actual variation in individual inventories may be incorrectly represented. Finally, the questionnaires do not yield relative frequencies of variable forms.

As remarked at the outset, 53% (= 161) of the total possible inventories were extracted. The inventories were chosen for geographical coverage but the regions are unevenly represented: the extracted grammars represent 97.3% of the West Midland counties speakers, 65.3% of the Southern and Northern counties and Man speakers, and only 35.6% of the East Midland counties speakers.

The data are limited to present-tense non-negated *be* inventories in basic, unmarked sentences. Inventories with positional variation are excluded, so only 39% of the above inventories are included. An inventory was assumed to have positional variation if the forms that appear in interrogative or focused positions before ellipsis are not a subset of the forms in uninverted, unstressed declarative positions.

(i) Excluded because of positional variation: *I'm thirsty. Be I right?*

(ii) Included: *I be thirsty. Be I right?*

(iii) Included: *I am/be thirsty. Be I right?*

(i) and (ii) are considered to be non-identical inventories. (i) is not covered by the constraint set of this study.

References

Anderwald, Lieselotte. 1999. *Negation in non-standard British English.* Freiburg: University of Freiburg Ph.D. dissertation.

Anderwald, Lieselotte. Forthcoming. *Was/Were* Variation in non-standard British English today. *English World-Wide.*

Anttila, Arto. 2000. Variation and phonological theory. To appear in Jack Chambers, Peter Trudgill, and Natalie Schilling-Estes (eds.), *Handbook of Language Variation and Change.* Oxford: Blackwell.

Anttila, Arto and Vivienne Fong. 2000. The partitive constraint in Optimality Theory. On-line, Rutgers Optimality Archive: http://ruccs.rutgers.edu/roa.html.

Asudeh, Arshia. 2001. Linking, optionality, and ambiguity in Marathi. To appear in Sells (ed).

Barron, Julia. 1998. Have contraction: explaining "trace effects" in a theory without movement. *Linguistics* 36:223–251.

Beckman, J., Dickey, L. and Urbanczyk, S. (eds), (1995). *Papers in Optimality Theory.* University of Massacusetts Occasional Papers 18. Amherst: University of Massachusetts.

Bender, Emily M. and Ivan A. Sag. In press. Incorporating contracted auxiliaries in English. In *Grammatical Interfaces in HPSG*, ed by Ronnie Cann, Claire Grover and Philip Miller. Stanford: CSLI Publications.

Benua, Laura. (1995) Identity effects in morphological truncation. In Beckman et al. (eds), 77–136.

Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences* 21: 43–58. University of Amsterdam. On-line, Rutgers Optimality Archive: http://ruccs.rutgers.edu/roa.html.

Boersma, Paul. 1998. *Functional Phonology [LOT International Series 11].* The Hague: Holland Academic Graphics. [Doctoral thesis, University of Amsterdam].

Boersma, Paul. 1999a. Optimality-Theoretic learning in the Praat program. *IFA Proceedings* 23: 17–35.

Boersma, Paul. 1999b. Review of Arto Anttila: *Variation in Finnish phonology and morphology. GLOT International* 5:1. To appear.

Boersma, Paul. 2000. Learning a grammar in Functional Phonology. In *Optimality Theory: Phonology, Syntax, and Acquisition*, ed. by Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer, 465–523. Oxford: Oxford University Press.

Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32: 45–86.

Boersma, Paul and David Weenink. 2000. Praat computer program. On-line, Institute of Phonetic Sciences, University of Amsterdam: http://www.fon.hum.uva.nl/praat/.

Bonet, Eulalia. 1995. Feature structure of Romance clitics. *Natural Language & Linguistic Theory* 13: 607–47.

Börjars, Kersti and Carol Chapman. 1998. Agreement and pro-drop in some dialects of English. *Linguistics* 36: 71–9.

Bresnan, Joan. 2000. Optimal syntax. In Dekkers et al. (eds), 334–85.

Bresnan, Joan. 2001a. Explaining morphosyntactic competition. In *Handbook of Contemporary Syntactic Theory*, ed. by Mark Baltin and Chris Collins, 11-44. Oxford: Blackwell Publishers.

Bresnan, Joan. 2001b. *Lexical-Functional Syntax.* Oxford: Blackwell Publishers.

Bresnan, Joan. In press a. The lexicon in Optimality Theory. In *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*, ed. by Suzanne Stevenson and Paola Merlo. Amsterdam: John Benjamins Publishers. On-line, Stanford University: http://www-lfg.stanford.edu/bresnan/download.html.

Bresnan, Joan. In press b. The emergence of the unmarked pronoun. In *Optimality-Theoretic Syntax*, ed. by Jane Grimshaw, Géraldine Legendre, and Sten Vikner. Cambridge, MA: The MIT Press.

Bresnan, Joan, Shipra Dingare, and Christopher Manning. 2001. Soft constraints mirror hard constraints: Voice and person in English and Lummi. *LFG 2001.*

Brown, Penelope and Steven Levinson. 1978/1987. *Politeness. Some Universals in Language Usage.* 2nd ed. Cambridge: Cambridge University Press.

Cheshire, Jenny. 1991. Variation in the use of *ain't* in an urban British English dialect. In Trudgill and Chambers (eds), 54–73.

Cheshire, Jenny. 1996. Syntactic variation and the concept of prominence. *Speech Past and Present. Essays in English Dialectology in Memory of Ossi Ihalainen*, 1–17. Frankfurt: Peter Lang.

Cheshire, Jenny, Viv Edwards, and Pamela Whittle. 1989. Urban British dialect grammar: the question fo dialect levelling. *English World-Wide* 10, 185–225.

Cheshire, Jenny, Viv Edwards, and Pamela Whittle. 1993. Non-standard English and dialect levelling. In Milroy and Milroy (eds), 53–96.

Clark, Brady. 2001. Modeling the constant rate of change in stochastic Optimality Theory. Stanford, CA: Stanford University Department of Linguistics unpublished MS.

Dekkers, Joost, Frank van der Leeuw, and Jeroen van de Weijer. (eds). 2000. *Optimality Theory: Phonology, Syntax, and Acquisition.* Oxford: Oxford University Press.

Dingare, Shipra. 2001. Effects of the definiteness hierarchy on rates of passivization in the WSJ corpus. Stanford, CA: Stanford University Department of Linguistics unpublished MS.

Foley, William. 1986. *The Papuan Languages of New Guinea.* Cambridge: Cambridge Uniersity Press.

Grimshaw, Jane. 1997a. Projection, heads, and optimality. *Linguistic Inquiry* 28: 373–422.

Grimshaw, Jane. 1997b. The best clitic: Constraint conflict in morphosyntax. In Liliane Haegeman (ed), *Elements of Grammar*, 169–96. Dordrecht: Kluwer Academic Publishers.

Grimshaw, Jane. In press. Optimal clitic positions and the lexicon in Romance clitic systems. In Grimshaw et al. (eds). Available on the Rutgers Optimality Archive: (ROA-374-01100)

Grimshaw, Jane, Géraldine Legendre, and Sten Vikner (eds.) In press. *Optimality-theoretic Syntax.* Cambridge: The MIT Press.

Guy, Gregory. 1997. Violable is variable: OT and linguistic variation. *Language Variation and Change* 9: 333–24.

Hayes, Bruce, Bruce Tesar, and Kie Zuraw. 2000. "OTSoft" software package. On-line, UCLA: http://www.linguistics.ucla.edu/people/hayes/otsoft/.

Hibiya, Junko. 2000. Variation studies and linguistic theory. Paper presented at the Eighteenth National Conference of the English Linguistic Society of Japan, November 18–19, 2000, Konan University, Kobe, Japan.

Hinskens, Frans, Roeland van Hout, and W. Leo Wetzels. 1997. *Variation, Change and Phonological Theory.* Amsterdam: John Benjamins.

Ihalainen, Ossi. 1991. On grammatical diffusion in Somerset folk speech. In Trudgill and Chambers (eds), 104–19.

Johnson, Mark. 1988. *Attribute-Value Logic and the Theory of Grammar.* Stanford: CSLI Publications.

Koontz-Garboden, Andrew. 2001. A stochastic OT approach to word order variation in Korlai Portuguese. *CLS 37.*

Kroch, Anthony S. 2001. Syntactic change. In *The Handbook of Contemporary Syntactic Theory*, ed. by Mark Baltin and Chris Collins, 699–729. Oxford: Blackwell Publishers.

Kuhn, Jonas. 1999a. Towards a simple architecture for the structure-function mapping. In *The Proceedings of the LFG '99 Conference, University of Manchester,* ed. by Miriam Butt and Tracy Holloway King. Stanford: CSLI Publications. On-line, CSLI Publications: http://csli-publications.stanford.edu/LFG/4/lfg99.html.

Kuhn, Jonas. 1999b. Generation and parsing in Optimality Theoretic syntax. Issues in the formalization of OT-LFG. To appear in Sells (ed.). On-line, University of Stuttgart: http://www.ims.uni-stuttgart.de/∼ jonas/.

Kuhn, Jonas. 2000. Faithfulness violations and bidirectional optimization. *Proceedings of LFG2000.* On-line, CSLI Publications: http://csli-publications.stanford.edu/LFG/5/lfg00.html.

Kuhn, Jonas. 2001. Formal and Computational Aspects of Optimality-theoretic Syntax. Stuttgart: IMS, Universität Stuttgart Ph.D. dissertation draft.

Kusters, Wouter. 2000. Morphological simplification: More than erosion? *Languages in Contact*, ed. by D. G. Gilbers, J. Nerbonne, and J. Schaeken (= *Studies in Slavic and General Linguistics*, vol. 28, 225–230. Amsterdam – Atlanta, GA: Rodopi.

Kusters, Wouter. 2001. Scandinavian and optimality theory. Leiden: Leiden University MS.

Lee, Hanjung. 1999. Discourse competing with syntax: "Misplaced" *que* in child French. Revised version of the paper presented at the Stanford/CSLI Workshop on Optimality Theory: Is Syntax Different? Common Cognitive Structures for Syntax and Phonology in Optimality Theory, Stanford, December 12-13, 1998, and at the 1999 Annual Meeting of the Linguistic Society of America, Los Angeles, January 7-10, 1999. On-line, Stanford University: http://www.stanford.edu/~hanjung/.

Lee, Hanjung. 2000. Bidirectional optimality and ambiguity in argument expression. A revised, expanded version of the paper presented at the LFG 2000 Conference, Berkeley, July 19-20, 2000. On-line, Stanford University: http://www.stanford.edu/~hanjung/.

Legendre, Géraldine, William Raymond, and Paul Smolensky. 1993. An Optimality-Theoretic typology of case and grammatical voice systems. *Proceedings of the Nineteenth Annual Meeting of the Berkeley Linguistics Society*, 464–78. Berkeley, CA.

Milroy, James and Lesley Milroy. 1993. *Real English. The Grammar of English Dialects in the British Isles.* London: Longman.

Nagy, Naomi and David Heap. 1998. Francoprovenal Null Subject and Constraint Interaction. *CLS 34: The Panels*, 151–66. Chicago: Chicago Linguistic Society.

Nordlinger, Rachel. 1998. *Constructive Case: Evidence from Australian Languages.* Stanford: CSLI.

Nordlinger, Rachel and Louisa Sadler. 2000. Tense as a nominal category. To appear in *Proceedings of LFG2000.* On-line, CSLI Publications: http://csli-publications.stanford.edu/LFG/5/lfg00.html.

van Oostendorp, Marc. 1997. Style levels in conflict resolution. In Hinskens et al. 207–29.

Orton, Harold et al. (eds). 1962–71. *Survey of English Dialects.* Leeds: Leeds, published for the University of Leeds by E. J. Arnold.

Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: constraint interaction in generative grammar.* RuCCS Technical Report #2. Piscateway, NJ: Rutgers University Center for Cognitive Science.

Schilling-Estes, Natalie and Walt Wolfram. 1994. Convergent explanation and alternative regularization patterns: *Were/weren't* leveling in a vernacular English variety. *Language Variation and Change* 6: 273–302.

Sadler, Louisa. 1998. English auxiliaries as tense inflections. *Essex Research Reports.*

Sells, Peter (ed). To appear. *Formal and Empirical Issues in Optimality-Theoretic Syntax.* Stanford, CA: CSLI Publications.

Smolensky, Paul. 1998. Why syntax is different (but not really). Paper presented at the Workshop "Is Syntax Different? Common cognitive structures for syntax and phonology in Optimality Theory," Center for the Study of Language and Information, Stanford University, December 12–13, 1998.

Spencer, Andrew. 1991. *Morphological Theory*. Oxford: Blackwell Publishers.

Trudgill, Peter. 1990. *The Dialects of England*. Oxford: Blackwell Publishers.

Trudgill, Peter and J. K. Chambers (eds). 1991. *Dialects of English. Studies in Grammatical Variation.* London: Longman.

Urbanczyk, Suzanne. 1995. Double reduplications in parallel. In Beckman et al. (eds), 499–531.