Optimality-theoretic Lexical Functional Grammar

Mark Johnson[*]
Cognitive and Linguistic Sciences, Box 1978
Brown University

1   Introduction

In her chapter in this volume, Bresnan (1998) describes a version of Lexical-Functional Grammar (LFG) in which Optimality Theory (OT) constraint satisfaction is used to identify well-formed linguistic structures. Bresnan shows how re-ranking of constraints changes the set of optimal outputs (surface forms), and uses this to elegantly account for a range of dialectal and cross-linguistic variation in the English auxiliary system. Bresnan's analysis has broad implications not just for the analysis of the auxiliary system, but for LFG and the study of parsing.

It should be clear that Bresnan's approach is new, and many of the details, both linguistic and formal, still remain to be worked out. In evaluating the broader implications of her work, I sometimes need to make assumptions about how these details will eventually be resolved. As in any theory whose foundations are still actively under development, formal properties and linguistic consequences may change dramatically as the theory is developed.

This paper begins with a review of Bresnan's proposals concerning inflectional systems, and evaluates the extent to which they make empirically-testable predictions. While Bresnan's current proposal is not yet precise enough to enable one to determine just what the range of possible agreement systems is predicted to be, it does seem that several putative linguistic universals concerning agreement can be stated in her framework.

Section 3 discusses the revision entailed by Bresnan's OT approach to the version of LFG presented in Kaplan and Bresnan (1982) and Kaplan (1995), called "classical LFG" below. Because Bresnan's analysis requires competition between *syntactic* analyses, it seems that her proposal adds a substantially new component to classical LFG. In this section I speculate that this added machinery may be able to supplant some of the constraint-based "unification" machinery of classical LFG.

Section 4 turns to issues of parsing with OT grammars of the general kind that Bresnan proposes. (Here "parsing" is used in its standard sense of "identifying the syntactic structure of a string", rather than the technical sense used in OT). The parsing problem OT LFG may turn out to be undecidable even though the corresponding problem for classical LFG is decidable. However it is not clear how relevant such a result would be, since the OT perspective itself suggests an alternative account in which sentence comprehension does not involve determining the grammaticality of the sentence being understood. This approach is conceptually related to maximum likelihood parsing, and suggests that it is closely related to a certain kind of probabilistic language model.

2   Bresnan's analysis of auxiliary selection

In classical LFG auxiliary selection is intimately related to agreement. Specifically, in a language with subject-verb agreement the lexical entries of tensed auxiliaries and verbs constrain the values of the subject's person and number features, so that each inflected lexical entry determines the range of subject agreement features it can appear with (Kaplan and Bresnan, 1982). This account had the advantage of formal simplicity, but the disadvantage that important substantive properties of the auxiliary system do not follow from the analysis. For example, every verb can appear with the full range of subject person and number features (i.e., there are no inflectional gaps), and by and large there is exactly one inflected form of each verb that agrees with each set of person and number features. This pattern follows from the architecture of Chomsky's original account of the English auxiliary system (Chomsky, 1957), but in classical LFG and similar unification-based theories it was an essentially accidental property of the lexicon as a whole.

As LFG developed the organization of the lexicon came to play a more prominent role, and facts of the kind just mentioned were captured via a more richly structured lexicon. For example, Andrews' Morphological Blocking Principle (1982; 1990) prevents insertion of a lexical item if another

2

lexical item from the same paradigm imposing more specific constraints could also be inserted. Thus an inflected form with no agreement constraints (i.e., a "default" form) is blocked if a more specialized form is present in the lexicon. Bresnan's (1998) OT account can be viewed as a radical extension of the Morphological Blocking Principle to all of syntax. Bresnan's examples in which the realization of verbal negation alternates between an inflectional element cliticized to an auxiliary and an independent lexical item (e.g., *aren't* vs. *are not*) provide evidence for competition at the syntactic as well as the morphological level.

Similar points are made by Grimshaw (1997), Legendre (to appear) and others. Indeed, it seems that Bresnan's analysis of lexical selection does not depend heavily on details of LFG, and could easily be re-expressed in a non-LFG OT framework. Perhaps this is because Bresnan's account follows primarily from the particular constraints she posits and their ranking, which do not depend on LFG-specific syntactic representations. The "unification-based" machinery of classical LFG seems largely superfluous to her analysis.

2.1   Inflectional Classes

In morphological blocking accounts such as Andrews', competing lexical forms are always ranked in terms of featural specialization, while in OT a language-particular constraint ranking determines how competing forms are ordered. Depending on the constraint-ranking, it may turn out that some candidate feature combinations are not the optimal surface forms for any input feature specification, so the constraint ranking effectively determines the range of possible candidate features and hence possible lexical entries.

For example, as Bresnan shows the presence of exactly the two specialized forms *am* and *is* in the present tense paradigm for BE follows from constraint ranking $*2, *PL \gg PARSE^{PERS\&NUM} \gg *1, *3, *SG$. However, note that the regular present tense verb paradigm contains only one specialized form (3rd singular), which would require a different constraint ranking, namely $*1, *2, *PL \gg PARSE^{PERS\&NUM} \gg *3, *SG$. Thus each inflectional class must be somehow associated with its own constraint ranking, rather than there being a single constraint ranking holding across a language.

Further, inflectional form selection in Bresnan's account seems to be fundamentally a choice between either a form that is specialized for a particular combination of input features or a general unspecialized form. However, not all inflectional patterns can be described in this way. For example, the more specialized form *was* surfaces in both the first and third person singular

forms of the past tense of BE in Standard English. The constraint ranking for present tense BE given by Bresnan would permit specialized forms to appear in these two positions in the paradigm, but does not explain their homophony.

2.2   Universals in OT LFG

Moving to more general issues, it is interesting to ask whether and how the OT LFG framework Bresnan outlines is capable of expressing putative typological universals that have been proposed elsewhere. Greenberg (1966) proposes several well-known universals concerning agreement. Some of these can be straight-forwardly expressed in Bresnan's framework, although they do not seem to follow from deeper principles.

> *Greenberg's Universal 32:*
> Whenever the verb agrees with a nominal subject or nominal object in gender it also agrees in number.

This could be expressed as a substantive universal requirement that every constraint ranking must satisfy, viz.:

> If g is a gender feature and PARSE^g $\gg$ *g, then there is a number feature n such that PARSE^n $\gg$ *n.

However, other universals proposed by Greenberg cannot be expressed so straight-forwardly.

> *Greenberg's Universal 37:*
> A language never has more gender categories in nonsingular numbers than it does in the singular.

This universal does not seem to be easy to express as a condition on constraint rankings, although sufficient conditions which ensure that the language generated by a constraint ranking satisfies this universal seem easy to state. For example, if *SG $\not\gg$ *PL then singular forms will never be more marked than corresponding plural forms, from which Universal 37 follows.

In any case, it is worth considering whether universals of the kind above should follow from innate principles of universal grammar, or should receive a different kind of explanation. Poverty of the stimulus arguments are the standard justification for assuming that linguistic universals follow from an innate universal grammar. But it is not clear how or even if they apply in the case of inflectional paradigms, since inflected forms are

directly observable. (Bresnan's account merely specifies the range of inflected forms a language may have, and says nothing directly about their phonological realization or the agreement relationships that they may enter into). Bresnan assumes that the sets of inputs and candidates are the same for all languages, i.e., universal, which means that any inflectional feature that appears in any language appears in the input. It seems then that the inputs contain inflectional features irrelevant to any given language. Of course it is possible to formulate a system in which inputs are universal but where inputs containing features inappropriate to the language concerned are ignored somehow (e.g., by associating them with a null surface form); such a system technically possesses a universal input set, but at the price of making the universal input set hypothesis essentially vacuous. Note that the lexicon and the candidate set must be language particular, since both include language-specific phonological forms.

3   Formal implications for LFG

The previous section focussed on the empirical implications of Bresnan's analysis. This section investigates the impact of Bresnan's adoption of OT competitive constraint satisfaction on the formal basis of LFG. Classical LFG as formulated in Kaplan and Bresnan (1982) and Kaplan (1995) is often described as a "constraint-based" theory of grammar (Shieber, 1992). The constraints in classical LFG are "hard" in the sense that a single constraint violation leads to ungrammaticality. Competition plays no role in classical LFG, although there have been other proposals besides Bresnan's to add it to LFG such as Frank et al. (1998).

What is the minimal modification to classical LFG one could make in order to make it compatible with Bresnan's account? For example, can Bresnan's account be regarded merely as a theory of the lexicon, where the lexical entries interact syntactically solely via the "hard" constraint mechanisms of classical LFG? It seems not, since one of the major points in Bresnan's paper is that competition between ranked constraints determines a language's multi-word syntactic constructions in the same way as it determines the language's lexical inventory. Thus OT competition cannot be restricted to the lexicon, and syntactic structures must be permitted to compete. This makes the mechanisms operative in the lexicon and the syntax much more uniform in Bresnan's account than they were in earlier LFG accounts. But as subsection 4.1 discusses, such syntactic competition may make the parsing problem much harder.

3.1 Feature structure constraints in OT LFG

In Bresnan's fragment the features associated with a candidate f-structure
are merely sets of atoms, while in a classical LFG the corresponding f-
structure would consist of attribute-value pairs. For example, Bresnan's
lexical entry for the candidate form *am* is merely [BE 1 SG], whereas in a
comparable classical LFG lexical entry each atom would appear as the value
of a unique attribute or f-structure function, i.e., [[PRED = BE], [PERSON =
1], [NUMBER = SG]]. This additional structure is necessary in classical LFG
and other "unification-based" theories since they rely on *functional unique-
ness* in their account of agreement. Agreeing elements both specify values
for the same attributes. Functional uniqueness requires that each attribute
have a single value, so if the values specified by the agreeing elements dif-
fer then the construction is ill-formed. For example, in a language with
subject-verb number agreement a singular subject specifies that the value
of its NUMBER attribute is SG, and a plural verb specifies that the value
of that same NUMBER attribute is PL. But the functional uniqueness con-
straint requires that the NUMBER attribute have a single value, so any
syntactic structure in which a singular subject appears with a plural verb
would be ungrammatical.

More abstractly, one role of the attributes in classical LFG is to for-
mally identify which feature values clash. Continuing with the example, SG
and PL clash because both are the value of the same NUMBER attribute,
while SG and 1 do not clash because they are values of different attributes.
Bresnan's use of atomic feature values rather than attribute-value pairs re-
flects the fact that functional uniqueness feature clashes play no role in her
account.

Bresnan's account focuses on the possible realizations of inflectional
forms within verb phrases, and does not discuss subject-verb agreement
per se. As far as I can tell, it is consistent with Bresnan's account to re-
gard her use of atomic features as abbreviations for attribute-value pairs
(the resulting f-structures seem to meet all of the conditions classical LFG
imposes on well-formed f-structures), and to use functional uniqueness to
force subject-verb agreement as sketched above. However, whenever a new
mechanism (in this case, OT competition between syntactic structures) is
added to the formal machinery of a theory, one should ask if that mechanism
supplants or makes redundant other mechanisms used in the theory.

Besides its role in agreement, functional uniqueness is also used in clas-
sical LFG to ensure that the grammatical functions of a clause (e.g., SUBJ,

OBJ, etc.) are not doubly filled. But recent work semantic interpretation in LFG has adopted a "resource-based" linear logic approach which enforces both functional completeness and functional uniqueness as a by-product of semantic interpretation (Dalrymple, 1999). Johnson (1999) extends this approach to provide a feature structure system without any functional uniqueness constraint.

Indeed, a direct extension of Bresnan's own analysis can account for subject-verb agreement without appealing to functional uniqueness. In this extension I distinguish the subject's semantic argument-structure features appearing in the input, which I write as 'SG', '1', etc., from the corresponding superficial verbal inflection features 'SG$_V$', '1$_V$', etc, which I take to appear in candidate representations only. (Presumably nominal inflection is encoded using similiar nominal features 'SG$_N$', '1$_N$', although for simplicity I ignore this here). The faithfulness constraint PARSE ensures that the input features appear in the candidates. I posit an additional constraint AGR$_S$, which is violated by a candidate representation whenever a verb's person or number inflection feature differs from its subject's corresponding feature in that candidate.[1] The ranking of the AGR$_S$ constraint relative to the constraints *SG$_V$ and *1$_V$ determines the possible inflected forms of a verb in exactly the same way that the relative ranking of PARSE$^{PERS\&NUM}$, *SG and *1 determines the inflected forms in Bresnan's account (presumably object agreement inflection is determined by the relative ranking of a similiar AGR$_O$ constraint).

Consider the example *I am*. Bresnan's analysis of present-tense *be*, expressed in terms of the constraints just discussed, corresponds to the constraint order PARSE $\gg$ *2$_V$, *PL$_V$ $\gg$ AGR$_S$ $\gg$ *1$_V$, *3$_V$, *SG$_V$. Just as in Bresnan's analysis, PARSE is a faithfulness constraint which is violated when an argument structure feature in the input fails to appear in a candidate: it appears undominated here because its role in Bresnan's analysis is played by AGR$_S$ here.

Input: [BE, SUBJ [PRO, 1, SG]]	PARSE	*PL$_V$, *2$_V$	AGR$_S$	*SG$_V$, *1$_V$, *3$_V$
'I are': [BE, SUBJ [PRO, 1, SG]]			*!*	
$\Rightarrow$ 'I am': [BE, 1$_V$, SG$_V$, SUBJ [PRO, 1, SG]]				**
'I is': [BE, 3$_V$, SG$_V$, SUBJ [PRO, 1, SG]]			*!	**
'She is': [BE, 3$_V$, SG$_V$, SUBJ [PRO, 3, SG]]	*!			**

[1]Because AGR^S only refers to candidate representations, the account does not require that agreement features appear in the input. In this approach it is not necessary to assume that language-specific agreement features appear in the input.

It should be clear that because of the close correspondence between this approach and Bresnan's, all of Bresnan's analyses can be expressed in the manner just described. Thus using just the mechanisms Bresnan assumes, it is possible to account for subject-verb agreement without appealing to f-structure constraints such as functional uniqueness. This raises the possibility that feature structure well-formedness constraints that play such a central role in classical LFG are not needed in OT LFG, leading to a radical simplification of the formal machinery of LFG.

4 Parsing in OT LFG

In computational linguistics and psycholinguistics, parsing refers to the identification of the syntactic structure of a sentence from its phonological string. In OT LFG, the universal parsing problem might be reasonably defined as follows:

> *The universal parsing problem for OT LFG:*
> Given a phonological string s and an OT LFG G as input, return the input-candidate pairs $\langle i, c \rangle$ generated by G such that the candidate c has phonological string s and c is the optimal output for i with respect to the ordered constraints defined in G.

The corresponding universal parsing problems for classical LFG and other unification-based theories are computationally difficult (NP-hard) but decidable (Barton, Berwick, and Ristad, 1987).

4.1 Complexity of OT LFG parsing

One might suspect that the global optimization over syntactic structures involved in OT LFG and other optimality-theoretic grammars may make their parsing problems more difficult than than those of corresponding theories without OT-style constraint optimization. This is because the well-formedness of a candidate representation may involve a comparison with candidates whose phonological strings differ arbitrarily from the string being analyzed. Just because a candidate is higher ranked than all other candidates with the phonological string being parsed does not guarantee that it is the optimal candidate for any input, since there may be higher ranked candidates with other phonological strings. The situation is depicted abstractly in Figure 1. In this figure the phonological string s_2 appears in two candidates c_2 and c_3. However, the input-candidate pair $\langle i_1, c_2 \rangle$ is not an optimal candidate since the pair $\langle i_1, c_1 \rangle$ is more optimal. On the other
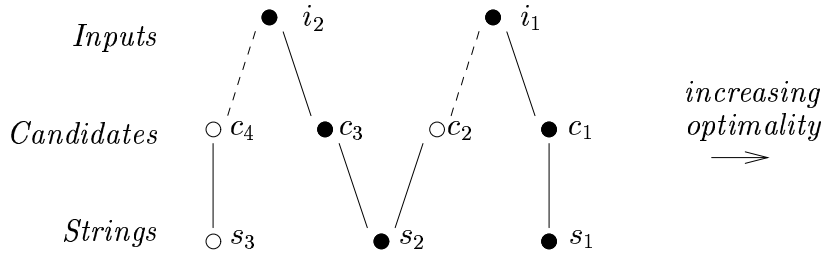
Figure 1: The highest ranked candidate (c_2) with a given phonological string (s_2) need not be an optimal candidate for any input, and an optimal candidate (c_3) for some input (i_2) need not be the highest ranked candidate for any string.

hand, the pair $\langle i_2, c_3 \rangle$ is optimal, even though the corresponding candidate c_3 is ranked lower than c_2. The phonological string s_3 is ungrammatical, since c_4, the only candidate with string s_3, is not the optimal candidate for any input.

In the mathematical study of parsing complexity it is standard to work with a simplification of the parsing problem called the recognition problem. The corresponding version of the universal recognition problem for OT LFG is:

> *The universal recognition problem for OT LFG:*
> Given a phonological string s and an OT LFG G, answer 'yes' if there is an input i which has an optimal candidate with s as its phonological string, otherwise answer 'no'.

Because a solution to the universal parsing problem implies a solution to the universal recognition problem, the complexity of the universal recognition problem is a lower bound on the complexity of the universal parsing problem. Depending on exactly how OT LFG is ultimately formalized, it may be possible to show that the universal recognition problem for OT LFG, and hence the universal parsing problem, is undecidable. The idea is to reduce the universal recognition problem for OT LFG to the emptiness problem for classical LFG, which is known to be undecidable (Kaplan and Bresnan, 1982). It is well-known that for any Turing machine M there is a classical LFG G_M whose terminal strings are precisely the sequences of moves of M's halting computations (Johnson, 1988). In effect, the string

w that G_M recognizes is the sequence of computational steps that the M performs. Using G_M to recognize w is equivalent to checking that w is in fact a legitimate sequence of computational steps for the machine M. This computation is not especially difficult, since w itself specifies exactly which steps must be checked. However, the problem of determining if any such w exists is considerably harder: indeed, there is no algorithm for determining if any such w exists, which means that the emptiness problem for classical LFG is undecidable.

The undecidability of the emptiness problem for classical LFG might be adapted to show the undecidability of the universal recognition problem for OT LFG as follows. Suppose that OT LFG is formalized in such a way that for every Turing machine M there is a grammar G'_M whose the candidate set consists of the set S_M of the syntactic structures generated by G_M plus a single extra syntactic structure s recognizable in some obvious way, say by having the unique terminal string "Doesn't Halt". (This is clearly possible if the candidate set in a OT LFG can be any set generated by a classical LFG). Further, suppose the constraints can be arranged so that every syntactic structure in S_M is more optimal than s. For example, one might introduce a feature FAIL which appears only on s, and introduce *FAIL as an undominated constraint. Then G'_M generates "Doesn't Halt" if and only if there are no syntactic structures more optimal than s, i.e., if and only if S_M is empty. But this latter condition holds if and only if the Turing machine M halts. Since there is no algorithm for determining if an arbitrary Turing machine M halts, there is also no algorithm for determining if the string "Doesn't Halt" is generated by G'_M, i.e., there is no algorithm which can solve the universal recognition problem if grammars such as G'_M are OT LFGs.

Thus the question becomes: under what assumptions would grammars such as G'_M be expressible as OT LFGs? Bresnan describes an OT LFG as having its input and candidate sets generated by classical LFGs. If no further constraints are imposed, then the procedure for constructing the classical LFGs G_M described in Johnson (1988) could be straight-forwardly adapted to generate OT LFGs G'_M as described above, and the undecidability result would presumably follow.

Would reasonable restrictions on OT LFGs rule out such pathological grammars? It is certainly true that construction just sketched yields grammars quite unlike linguistically plausible ones. But this observation does not justify ignoring such complexity results; rather it challenges us to try to

make precise exactly how the artificial grammars required for the complexity proof differ are linguistically implausible. Note that the construction makes no assumptions about the input set (indeed, it is systematically ignored), so assuming it to be universally specified has no effect on the construction. A restriction on the kinds of admissible candidate sets might rule out this kind of construction: the difficulty is precisely identifying linguistically plausible restrictions that do this.

4.2 Alternative perspectives on parsing

The undecidability argument sketched above requires that candidates with unboundedly differing structures compete, but in Bresnan's examples of OT LFGs the optimization involved seems to be strictly clause local, i.e., the global optimum can be obtained by optimizing each clause independently. Further, if there are only a finite number of clausal input feature combinations and candidate clausal structures then it may be possible to precompute for each lexical item the range of input clauses for which it appears in the optimal candidate. Under such conditions, OT LFG parsing need not involve an explicit optimization over candidates with alternative phonological strings, but might be "compiled" into a parsing process much like one for classical LFG. (Tesar (1995) exploits similiar locality properties in his "parsing" algorithm, while Frank and Satta (1998) and Karttunen (1998) show how a different kind of OT grammar can be compiled into a finite-state transducer). In such a system the OT constraints would serve to specify the morphosyntactic inventory of a language (i.e., account for cross-linguistic variation), but might not actually be used on-line during parsing.

A more radical approach is to reformulate the OT LFG parsing problem so that parsing only optimizes over candidates with the same phonological string, perhaps as follows:

> *The revised universal parsing problem for OT LFG:*
> Given a phonological string s and an OT LFG G (i.e., a set of ranked constraints and a lexicon), find the optimal candidates from the set of all candidates with s as their phonological string.

Under this revision, a parser presented with input s_2 in Figure 1 would produce c_2 as output, even though c_2 is not an optimal candidate for any input. This revised parsing problem could be computationally much simpler than the OT LFG parsing and recognition problems, as optimization over

candidates with phonological strings that differ arbitrarily from the string being parsed (a crucial component of the undecidability proof sketch just presented) no longer occurs. Frank et al. (1998) have extended a classical LFG parser in exactly this way. Stevenson and Smolensky (1997), working in a slightly different framework, show how this kind of model can account for a variety of psycholinguistic phenomena. They also point out that grammatical constraints may need to be reinterpreted or reformulated if they are to be used in such a parsing framework, and this seems to be true in the OT LFG setting as well. Indeed, it is not clear how or even if Bresnan's analysis could be restated in this framework.

Smolensky (1997) points out that in general the set of phonological forms generated by an OT grammar is a subset of the set of phonological forms which receive an analysis under the revised parsing problem above. It seems that the language generated by an OT LFG differs quite dramatically from the language accepted under the revised definition of the parsing problem. However, this may not be altogether bad, since humans often assign some interpretation to ungrammatical phonological strings. For example, the phonological string *I aren't tired* is interpretable, yet it is not the phonological string of any input's optimal candidates in Bresnan's OT LFG. Schematically, such a string may play the role depicted by s_3 in Figure 1; it is ungrammatical since it is not the optimal candidate for any input, but under the revised definition of the parsing problem it receives the parse c_4.

4.3   Optimality Theory and Probabilistic Grammars

Prince and Smolensky (1998) speculate that there is a "deep" relationship between optimality theory and connectionism. This section presents a related result, showing a close connection between the revised OT parsing problem and the maximum likelihood parsing problem, which is often adopted in probabilistic parsing. Both problems involve selecting a parse of the phonological string which is optimal on an ordinal scale, defined by ranked constraint violations in the case of OT, or a probability distribution in the case of probabilistic parsing.

Specifically, it seems that the revised OT parsing problem is closely related to a very general class of probabilistic models known as Gibbs distributions, Markov Random Fields models, or Maximum Entropy models. See Jelinek (1997) for an introduction, Abney (1997) for their application to constraint-based parsing, and Johnson et al. (1999) for a description of a stochastic version of LFG using such models. In this kind of model, the log-

arithm of the likelihood $P(\omega)$ of a parse ω is a linear function of real-valued properties $v_i(\omega)$ of the parse, i.e.,

$$P(\omega) \;=\; \frac{1}{Z} e^{\sum_{i=1,\ldots,n} -\lambda_i v_i(\omega)}.$$

In this class of models, $v_i(\omega)$ is the value of the ith of n properties of the parse ω, λ_i is an adjustable weight of property i, and Z is a normalization constant called the "partition function". The theory of these models imposes essentially no constraints on what the properties v_i can be, so we can take the properties to be the constraints of an OT grammar and let $v_i(\omega)$ be the number of times the ith constraint is violated by ω.

Suppose there is an upper bound c to the number of times any constraint is violated on any parse,[2] i.e., for all ω and i, $v_i(\omega) \leq c$. For simplicity assume that the OT constraint ranking is a linear order, i.e., that the ith constraint out-ranks the $i+1$th constraint. This implies that the OT parse ranking is the same as the lexicographic ordering of their property vectors $\tilde{v}(\omega)$. Set $\lambda_i = (c+1)^{n-i}$, which ensures that a single violation of the ith constraint will outweigh c violations of constraint $i+1$. It is straightforward to check that for all parses ω_1, ω_2, $P(\omega_1) > P(\omega_2)$ iff $\tilde{v}(\omega_1)$ lexicographically precedes $\tilde{v}(\omega_2)$, which in turn is true iff ω_1 is more optimal than ω_2 with respect to the constraints.

This result shows that if there is an upper bound on the number of times any constraint can be violated in a parse, the revised OT parsing problem can be reduced to the maximum likelihood parsing problem for a Gibbs form language model. It implies that although OT grammars are categorical (i.e., linguistic structures classified as either grammatical or ungrammatical), they are closely related to probabilistic language models; indeed, they are limiting cases of such models. This raises the possibility of applying techniques for parsing and learning for one kind of model to the other. For example, it might be interesting to compare the constraint re-ranking procedure for learning OT constraint rankings presented in Tesar and Smolensky (1998) with the statistical methods for estimating the parameters λ_i of a Gibbs distribution described in Abney (1997), Jelinek (1997) and Johnson et al. (1999).

---

[2]Frank and Satta (1998) and Karttunen (1998) also assume such a bound.

5 Conclusion

The Optimality-theoretic version of Lexical Functional Grammar that Bresnan (1998) provides not only an interesting account of cross-linguistic variation in the lexical inventories of auxiliary verbs and negation, it also provides a framework in which linguistic universals can be systematically explored. It has implications for the formal basis of LFG and other "unification-based" grammars, as it suggests that other linguistic processes, such as agreement, can be viewed in terms of competitive constraint satisfaction. Perhaps as importantly, by recasting LFG into a ranked constraint setting, Bresnan's work suggests novel ways of approaching parsing and learning in LFG. Specifically, the fact that well-formedness in Optimality Theory is defined in terms of an optimization suggests a close connection with probabilistic language models.

As noted above, Bresnan's analysis does not depend heavily on the details of LFG's syntactic representations, and it seems that it could be re-expressed in a variety of OT-based syntactic frameworks. Indeed, it is only necessary that we be able to identify the constraint violations Bresnan posits from the candidate structures; exactly how these constraint violations are encoded in candidate structures seems to be of secondary importance. This seems to be a general property of OT-based accounts. Thus from the perspective of both parsing and learning, the details of the representations used in an OT account are less important than the kinds of constraints that the account posits.

References

Abney, Steven P. 1997. Stochastic Attribute-Value Grammars. *Computational Linguistics*, 23(4):597–617.

Andrews, Avery. 1990. Unification and morphological blocking. *Natural Language and Linguistic Theory*, 8:507–557.

Andrews, Avery D. 1982. The representation of Case in modern Icelandic. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*. The MIT Press, Cambridge, Massachusetts, pages 427–502.

Barton, Jr., G. Edward, Robert C. Berwick, and Eric Sven Ristad. 1987. *Computational Complexity and Natural Language*. The MIT Press, Cambridge, Massachusetts.

14

Bresnan, Joan. 1998. The lexicon in optimality theory. In Paula Merlo and Susan Stevenson, editors, *The Proceedings of the 1998 CUNY conference.* ???, page ???

Chomsky, Noam. 1957. *Syntactic Structures.* Mouton, The Hague.

Dalrymple, Mary. 1999. *Semantics and Syntax in Lexical Functional Grammar: The Resource Logic Approach.* The MIT Press, Cambridge, Massachusetts.

Frank, Anette, Tracy Holloway King, Jonas Kuhn, and John Maxwell. 1998. Optimality Theory style constraint ranking in large-scale LFG grammars. In Miriam Butt and Tracy Holloway King, editors, *Proceedings of LFG'98*, Stanford, California. CSLI Press.

Frank, Robert and Giorgio Satta. 1998. Optimality Theory and the generative complexity of constraint violability. *Computational Linguistics*, 24(2):307–316.

Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg, editor, *Universals of Language.* The MIT Press, Cambridge, Massachusetts, chapter 5, pages 73–113.

Grimshaw, Jane. 1997. Projection, heads, and optimality. *Linguistic Inquiry*, 28(3):373–422.

Jelinek, Frederick. 1997. *Statistical Methods for Speech Recognition.* The MIT Press, Cambridge, Massachusetts.

Johnson, Mark. 1988. *Attribute Value Logic and The Theory of Grammar.* Number 16 in CSLI Lecture Notes Series. Chicago University Press.

Johnson, Mark. 1999. A resource sensitive interpretation of Lexical Functional Grammar. *The Journal of Logic, Language and Information*, 8(1).

Johnson, Mark, Stuart Geman, Stephen Canon, Zhiyi Chi, and Stefan Riezler. 1999. Estimators for stochastic "unification-based" grammars. In *The Proceedings of the 37th Annual Conference of the Association for Computational Linguistics*, Orlando, Florida. Harcourt Brace.

Kaplan, Ronald M. 1995. The formal architecture of LFG. In Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell III, and Annie Zaenen, editors, *Formal Issues in Lexical-Functional Grammar*, number 47 in CSLI Lecture Notes Series. CSLI Publications, chapter 1, pages 7–28.

Kaplan, Ronald M. and Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*. The MIT Press, chapter 4, pages 173–281.

Karttunen, Lauri. 1998. The proper treatment of optimality in computational phonology. In Lauri Karttunen and Kemal Oflazer, editors, *Proceedings of the International Workshop on Finite-State Methods in Natural Language Processing*, pages 1–12, Bilkent University, Ankara, Turkey.

Legendre, Géraldine. to appear. Morphological and prosodic alignment of bulgarian clitics. In J. Dekkers, F. van der Leeuw, and J. van de Weijer, editors, *Optimality Theory: Syntax, Phonology, and Acquisition*. Oxford University Press.

Prince, Alan and Paul Smolensky. 1998. *Optimality Theory*. The MIT Press, Cambridge, Massachusetts.

Shieber, Stuart M. 1992. *Constraint-based Grammar Formalisms*. The MIT Press, Cambridge, Massachusetts.

Smolensky, Paul. 1997. On the comprehension/production dilemma in child language. *Linguistic Inquiry*, 27(4):720–731.

Stevenson, Suzanne and Paul Smolensky. 1997. Optimal sentence processing. Paper presented at the Hopkins Optimality Theory Fest.

Tesar, Bruce. 1995. *Computational Optimality Theory*. Ph.D. thesis, University of Colorado.

Tesar, Bruce and Paul Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry*, 29(2):229–268.