

GRAMMATICALITY AND UNGRAMMATICALITY IN PHONOLOGY

ABSTRACT*

In this paper, I make two theoretical claims: (i) For some form to be grammatical in language *L*, it is not necessary that the form satisfy all constraints that are active in *L*, i.e. even grammatical forms can violate constraints. (ii) There are degrees of ungrammaticality, i.e. not all ungrammatical forms are equally ungrammatical. I first show that these claims follow straightforwardly from the basic architecture of an Optimality Theoretic grammar. I then show that the surface sound patterns used most widely in formal phonology cannot be used to test the truth of these two claims, but argue that results from speech processing experiments can. Finally, I discuss three experiments on the processing of non-words of the form [stVt], [skVk] and [spVp] in English that were designed to test these claims, and show that both claims are confirmed by the results of the experiments.

* This paper benefited greatly from discussion with and feedback from many people, and I express my appreciation to all of them: John McCarthy, John Kingston, Joe Pater, Pam Beddor, Sally Thomason, Jaye Padgett, the audiences at GLOW 27, LabPhon 9, the University of Maryland, Cornell University, Stanford University, Indiana University, Harvard University, the University of Massachusetts, and the University of Michigan. I also need to express my thanks to the anonymous reviewers, the associate editor in charge of this paper, and the editor of the journal. Their insistence on high standards, and their suggestions for improvements, resulted in this paper becoming much better. I, of course, accept all responsibility for the views expressed in the paper.

1. INTRODUCTION

As linguists, we develop grammatical models to explain some observed phenomenon of linguistic competence, and we deem a model successful if it can account for the phenomenon. However, the models we develop often make predictions beyond the data they were designed to account for. If these extrapolatory predictions are corroborated by additional data it is usually taken as evidence that the model is, if not correct, at least on the right track. In this paper, I point out two such extrapolatory predictions that follow from the basic architecture of an Optimality Theoretic grammar (Prince and Smolensky 1993/2004): (i) even grammatical forms violate some constraints (i.e. grammatical forms do not have to be perfect), and (ii) there are degrees of ungrammaticality (i.e. not all ungrammatical forms are equally bad).

Theoretical phonology relies primarily on ‘surface sound patterns’ (Ohala 1986) for evidence about phonological grammar. However, this kind of data is not effective in addressing all questions that are relevant to phonological theory. Pierrehumbert et al. (2000:274) argue that the increasing prominence of laboratory approaches to phonology is driven by the realization that there are ‘issues in phonology that are not effectively addressed using traditional types of data’. The two predictions mentioned in the previous paragraph are examples of such issues that cannot be resolved by relying only on surface sound patterns. This does not mean that these predictions are untestable, but only that we have to rely on other kinds of data. In this paper, I show how psycholinguistic experiments can be used to test these two predictions.

If the first prediction is true, we should see a processing difference between two grammatical forms that differ in terms of the constraints that they violate. A grammatical form that violates more, or more serious, constraints should show a processing deficit relative to a grammatical form that violates fewer, or less serious, constraints. Similarly, if the second prediction is true, a processing difference should be observed between ungrammatical forms that differ in terms of their degree of ungrammaticality. An ungrammatical form that violates more or more serious constraints should show a deficit relative to a form that violates fewer or less serious constraints. I use this logic to test these two predictions, showing that both predictions can be confirmed as correct.

The validity of this research program, as well as the interpretation of the results of the experiments, rests on several not uncontroversial assumptions. First, a realist stance has to be taken with regard to grammatical models. It has to be accepted that grammatical models such as Optimality Theory (henceforth OT) is more than just a convenient and succinct way to describe linguistic data and observations. Data from psycholinguistic processing experiments can only be relevant for grammatical models if such models are assumed to reflect the psychological reality of the language user’s linguistic competence. Secondly, many things other than formal grammar are known to influence language processing. Specifically, we know that usage-based frequencies have a strong influence on many aspects of processing (see section 3 below). Results from processing experiments can hence be interpreted as evidence about formal grammar only if it is assumed that both grammar and usage frequencies influence processing.

In this paper, I subscribe to these two assumptions. The first assumption is a philosophical stance that cannot be shown to be correct or incorrect. The second can more easily be tested empirically, and it is a topic of much recent research, as the discussion in section 3 shows.

In what follows, I first discuss the two predictions derived from OT in more detail, and then discuss the relationship between formal grammar and usage-based frequencies, reviewing recent literature that shows that grammar, though influenced by usage frequencies, is also independent from it. I then turn to a consideration of a restriction on English words of the form [sC_iVC_i], and develop a standard OT account of this restriction. This account provides typical examples of the two predictions noted above. The core of the paper follows, with a discussion of the results of three experiments designed to test the predictions illustrated in the OT account.

2. GRAMMATICALITY AND UNGRAMMATICALITY IN OPTIMALITY THEORY

2.1. WHAT IT MEANS TO BE GRAMMATICAL

In the earliest instantiations of generative phonology, constraints played a very small role or no role at all. The bulk of *The Sound Pattern of English* ('SPE', Chomsky and Halle 1968) is a purely rule-based derivational theory. However, it was realized very soon that a theory devoid of constraints was inadequate. The most important impetus for assigning constraints some role in phonology was the focus on phonological conspiracies in the 1970's. In an important paper, Kisseberth (1970) showed that several seemingly unrelated phonological rules of Yawelmani Yokuts conspire to achieve the same purpose – they all either actively eliminate or fail to create a sequence of three adjacent consonants. The unity between these rules can only be expressed in terms of a constraint against such sequences, i.e. *CCC. Ever since the discovery of conspiracies, much of generative phonology has been occupied with determining the nature of phonological constraints.

An assumption of nearly all of generative phonology until the advent of OT was that constraints are inviolable. Prince and Smolensky (1993:6) refer to the 'widely held view that constraints are statements of ... phonotactic truth'. McCarthy and Prince (1993:5) say about phonology before 1990: 'On the usual view, the output is the form which meets *all* the relevant constraints ...' [emphasis in the original]. This implies that a grammatical form in some language cannot violate any constraint that holds for that language. A grammatical form of language *L* is a form that is 'perfect' in terms of the constraints of *L*. A problem with this view of constraints is that, with very few exceptions, every constraint that has ever been proposed is violated in grammatical forms of at least some language.¹

Several different solutions to this problem were explored during the last quarter of the previous century. One approach was to reject any constraint that is violated in some language. See, for instance, Odden (1986) who questions the existence of the tonal Obligatory Contour Principle (OCP) based on the fact that there are many languages that tolerate adjacent high tones. Another approach was to reanalyze examples that seem to violate some constraint in such a way that they no longer do. This usually involved limiting the application of a constraint to a specific stage in the derivation. McCarthy's 1986 treatment of the OCP-Place constraint can serve as an example. McCarthy argues that the ban on adjacent homorganic consonants in the verbal root in

¹ The prohibition against crossing association lines in tonal phonology is an example of a constraint that is probably never violated (Sagey 1988). In Optimality Theory, it is sometimes assumed that such truly inviolable constraints are built into GEN so that forms that violate these constraints will not even be generated. However, see De Lacy (2007) for a different interpretation of these inviolable constraints.

Semitic languages is the result of a constraint against adjacent consonants at the same place of articulation (OCP-Place). The existence of verbs with identical adjacent consonants, like *galal* ‘he rolled/revolved’ in Hebrew, then presents a problem since these verbs seem to violate OCP-Place.² McCarthy’s solution is to reinterpret OCP-Place as a morpheme structure constraint, and to assume that a verb like *galal* is derived from a bi-consonantal root /g-l/. On the level at which OCP-Place applies, this verb hence does not have adjacent homorganic consonants so that it does not violate OCP-Place. A third approach was to treat constraints like parameters. Itô (1989) takes this approach in her treatment of the avoidance of onsetless syllables. She claims that there is an ‘absolute Onset Principle’ that forbids all onsetless syllables. This constraint is part of the grammar of a language like Arabic in which any form with an onsetless syllable is ungrammatical. However, since English allows onsetless syllables, this constraint is not part of the grammar of English at all. Under this approach constraints are still inviolable, but they are no longer universal.

OT represents a departure from all these previous approaches to constraints. In OT, constraints are assumed to be universal but violable. The fact that there are languages that tolerate adjacent high tones is hence not a reason to reject the tonal OCP. In languages like these, the tonal OCP is simply ranked low enough that violation of this constraint is tolerated. See, for instance, Myers (1997) for an OT-reanalysis of many of the examples that Odden (1986) claims as counterexamples to the OCP. The OCP is included in Myers’s analyses, but violation of the constraint is tolerated under domination of higher-ranking constraints. Berent, Everett and Shimron (2001) and Coetzee (2004, to appear) give OT-reanalyses of Hebrew verbs like *galal*. In their analyses, verbs like these do violate OCP-Place, but this violation is tolerated since this constraint is crucially dominated by higher ranking constraints. There are also many examples of OT-analyses in which grammatical forms violate ONSET, the constraint requiring syllables to have onsets.

From the preceding discussion, it is clear that what it means to be grammatical in language *L* depends amongst other things on whether constraints are taken to be inviolable. In a theory where constraints are inviolable, a form is grammatical in *L* only if it is perfectly well-formed. However, in OT this is not true. In OT, even grammatical forms may violate constraints. A grammatical form is a form with minimal constraint violation (as determined by the constraint ranking) and not a form with no constraint violations. Burzio (1995:5) remarks in this regard that grammaticality has been reconceptualized in OT and that it should no longer be defined in terms of well-formedness but rather in terms of ‘best’-formedness.

We are faced with two contradictory views about what it means to be grammatical. Traditional data, the ‘surface sound patterns’ of Ohala (1986), cannot adjudicate between these two views. For instance, we know that English allows words that violate the OCP-Place constraint that is active in, for instance, Hebrew – i.e. English has words with homorganic consonants such as *fame* (with two labials) and *keg* (with two dorsals). So, we know that the grammar of English should deem words with homorganic consonants to be grammatical. But the English grammar would do this if OCP-Place is not part of the grammar of English, and if OCP-Place is in the grammar of English but is freely violated. To decide between these two

² The two [l]’s are of course not adjacent in the surface form, but are separated by the vowel [a]. Following in the tradition of McCarthy (1985), I assume that Semitic consonants and vowels are represented on different plains. The two [l]’s are hence adjacent on the consonantal plain.

conceptualizations of what it means to be grammatical we have to rely on alternative data sources. In this paper, I show how the results of a speech processing experiment can be used to test these two views of grammaticality.

A clarifying note is called for here. There is evidence from ‘surface sound patterns’ for the activity of some constraints that are claimed to be violated by grammatical forms in typical OT-analyses. For instance, the English word *and* is claimed to violate ONSET, but to be grammatical nonetheless since ONSET is crucially dominated by the anti-epenthesis and anti-deletion constraints MAX and DEP. Of course, *and* would be deemed grammatical in a grammar with the ranking {MAX, DEP} >> ONSET, but also in a grammar that lacked ONSET altogether. This seems similar to the OCP-Place example in the previous paragraph, but it is not quite the same. When a consonant can be syllabified either into onset or coda position without earning a violation of a faithfulness constraint, the onset syllabification is preferred. Concretely, the word *demand* is syllabified as *de.mand* not as *dem.and*. This shows that, when faithfulness is not at stake, ONSET exerts its force even in English. This is a very common phonological phenomenon, and has been dubbed ‘the emergence of the unmarked’ (TETU) by McCarthy and Prince (1994, 2003).

Even though TETU gives evidence for the activity in English of a constraint that prefers syllables to have onsets, it still does not necessarily imply that a word like *and* violates this constraint. Itô’s 1989 analysis of syllable structure distinguishes two constraints on syllables, a strong version that requires all syllables to have onsets and a weak version that does not enforce onsetless syllables but favors them when possible. Under this view, the English grammar would contain only the weak onset constraint that would be violated by the syllabification *dem.and* for *demand*. But since the strong onset violation is not included in the grammar of English, *and* would not violate any constraint that references onsets. Further more, there are many constraints for which not even TETU gives any evidence. For instance, there is no evidence from TETU for the activity of OCP-Place in English. There is hence no evidence from surface sound patterns that words like *fame* and *keg* do violate an OCP-Place type constraint. Surface sound patterns cannot resolve all of these examples, and it is necessary to rely on other data sources such as psycholinguistic experimental evidence.

2.2. DEGREES OF UNGRAMMATICALITY

In most of the generative tradition, grammar is seen as a function that maps every input onto a unique grammatical output. The phonological grammar of some language *L* hence divides the space of potential utterances into two disjoint sets, grammatical forms (possible words) and ungrammatical forms (impossible words).³ But does grammar make finer distinctions within the set of ungrammatical forms? Or put differently, are there degrees of ungrammaticality?

In generative syntax there has always been at least an informal acceptance of the idea of degrees of ungrammaticality. Chomsky refers to this in *Aspects* (1965:149), implying that a

³ A form can be ruled out as ungrammatical for many different reasons. For instance, it can violate a word-minimality type constraint in which case it is truly an ‘impossible word’. But it can also be ruled out because it contains a disallowed sound sequence although it meets all other requirements for wordhood. A form can also be ruled out because of problems in its morphological structure – for instance, it may consist of bound root lacking any affix. The distinction between these different reasons for ungrammaticality is not relevant in this paper, and for the sake of simplicity, I will just refer to ungrammatical forms as ‘impossible words’.

sentence that violates two constraints of syntactic well-formedness is more ungrammatical than a sentence that violates only one such constraint (see also Chomsky 1986: section 7, and Epstein 1990, 2000). In generative phonology, degrees of ungrammaticality have received even less attention, the reason in all likelihood being the lack of a formal mechanism to derive degrees of ill-formedness from the grammar. More recent work in the laboratory phonology tradition has explored this concept, but the focus of this research has been more on how usage frequency statistics calculated over the lexicon influence word-likeness (see section 3 below for more on this).

The architecture of an OT grammar presents a way in which degrees of ungrammaticality can be derived in a straightforward manner. In an OT grammar, some form is deemed an impossible word if (i) it violates some markedness constraint M , and (ii) if M outranks some faithfulness constraint F that can be violated to avoid violation of M . In English, for example, [bdif] is not a possible word since it has a sonority plateau in the onset of the first syllable – a structure that English does not tolerate (though many other languages, like Russian, do – cf. Russian *bdenie* ‘vigil’). In OT, the ungrammaticality of [bdif] can be captured by ranking a constraint against sonority plateaus, *PLATEAU, above some faithfulness constraint that can be violated to avoid violation of *PLATEAU. If *PLATEAU outranks the anti-epenthesis constraint DEP, an input like /bdif/ can never be mapped onto itself but will always surface with an epenthetic vowel, i.e. [bədif] (see Berent et al. 2007 for evidence that English uses schwa epenthesis to repair such sequences, and Davidson 2006 for a different view). But English also does not allow sonority falls in syllable onsets; that is a form like [lbif] is also ungrammatical, and an input like /lbif/ would also be realized with an epenthetic vowel as [ləbif]. (Languages like Russian, again, do tolerate sonority falls in onset position – cf. Russian *lba* ‘forehead-GEN’). This can be captured by ranking a constraint against sonority falls, *FALL, also above DEP.

It is clear that both *PLATEAU and *FALL rule out forms as ungrammatical, and both rank above DEP. If *PLATEAU and *FALL are unranked relative to each other, then [lbif] and [bdif] are equally ungrammatical. However, if there is a ranking between *PLATEAU and *FALL, then it would become possible to derive an ill-formedness difference between [lbif] and [bdif]. The one that is ruled out by the higher ranking constraint would in some sense be worse. Berent et al. (2007) show that languages that tolerate violation of *FALL typically also allow violation of *PLATEAU (see also Greenberg 1978). Following in a tradition that started with Prince and Smolensky (1993/2004), such implicational universals are captured in OT by imposing a fixed, universal ranking between constraints, in the case at hand, *FALL >> *PLATEAU. This gives the ranking *FALL >> *PLATEAU >> DEP for English. Under this ranking, both [bdif] and [lbif] are ungrammatical, since both of them violate a markedness constraint that outranks the faithfulness constraint DEP. However, since [lbif] violates the higher-ranking markedness constraint, there is a sense in which it is worse than [bdif].

Since the markedness constraints that rule out impossible words can be ranked relative to each other, an OT grammar generates information about the ill-formedness relationship between different ungrammatical forms. It should be noted that although this information is generated and can hence in principle be used to impose an ill-formedness scale on ungrammatical forms, it does not have to be used. In fact, some OT practitioners are of the opinion that this information, though available, is irrelevant and has no real meaning. Legendre et al. argue specifically that all an OT grammar does is to select grammatical forms, and that it does not have anything to say about the relationship among ungrammatical forms: ‘the highest-Harmony member of a

candidate set is the only grammatical one, all other candidates being equally ungrammatical' (2006:353, see also Tesar and Smolensky 1998:230). If this stance is taken, then an OT grammar generates information about the relation between ungrammatical forms that is irrelevant and hence ignored. An OT grammar could then be viewed as more powerful than necessary since it generates irrelevant information. An alternative view is that an OT grammar does more than just select the grammatical forms, and that it does actually use the information generated about ungrammatical forms to impose an ill-formedness hierarchy on these forms. See Berent, Everett and Shimron (2001) and Silva (2004) for examples of this approach, and Coetzee (2004, 2006) for a formal development.

The data traditionally used in theoretical linguistics, surface sound patterns, cannot be used to test whether there are degrees of ungrammaticality. To continue with the example from above, all that this kind of data can show is that there are no English words with onset clusters like [bd-] or [lb-]. These data are silent about any well-formedness difference between these two kinds of forms in English. To test the claim that there are degrees of ungrammaticality, we therefore again have to rely on other kinds of data. In this paper, I show how psycholinguistic experiments can be used to investigate this question.

3. ON THE RELATION BETWEEN GRAMMAR AND FREQUENCY

One of the assumptions of this paper is that grammar influences linguistic performance, and that performance data can therefore serve as a window on the internal workings of grammar. By grammar, I refer here to formal abstract grammatical principles such as those expressed in a typical OT analysis. Although these grammatical principles are certainly influenced by experience – after all, the grammar of some language can only be learned by exposure to the language – I consider them to be more than and partially independent from generalizations over language usage frequency statistics. This view stands in contrast to a purely usage-based approach to grammar: 'While all linguists are likely to agree that grammar is the cognitive organization of language, a usage-based theorist would make the more specific proposal that grammar is the cognitive organization of one's experience with language' (Bybee 2006:711). If this view is taken to its extreme, then 'phonological grammar is a simple projection of lexical statistics' (Hay et al. 2003:59). I argue in sections 5, 6 and 7 that the results of psycholinguistic experiments that I discuss there cannot be explained solely on the basis of usage frequency statistics calculated over the lexicon. In that section, I motivate this claim with regard to the specific experiments that I discuss there. Here, however, I briefly review some of the recent literature that supports the partial independence of formal grammar from usage frequency. I review three kinds of evidence: (i) Language users have linguistic knowledge about structures with which they have no experience. (ii) Successful grammar learning requires some prior linguistic knowledge – i.e. knowledge that does not depend on experience. (iii) Not all results of speech processing experiments can be explained by reference to usage frequencies.

In a recent study, Berent et al. (2007) presented English listeners with non-word stimuli with initial clusters, all of which are unattested in English: with a sonority fall [lbif], a sonority plateau [bdif], or a small sonority rise [bnif]. Dupoux et al. (1999, 2001) have shown that Japanese listeners tend to perceive vowels between the members of consonant clusters that are

unattested in their language.⁴ Berent et al. found the same effect with the English listeners in their study. They presented the tokens to their subjects in a syllable count task, so that for each token, subjects had to say whether it has one syllable (implying that they did not perceive an extra vowel between the consonants), or two syllables (implying that they did perceive an extra vowel). Subjects gave two syllable responses more than half the time. If usage frequency is solely responsible for this perceptual epenthesis effect, then it should be equally likely to occur in each of the three cluster types, since subjects have equal experience with all three, namely none. This is not what Berent et al. found. Their subjects were most likely to perceive an epenthetic vowel in clusters with a sonority fall [lbif], next most likely in clusters with a plateau [bdif], and least likely in clusters with a small rise [bnif]. The likelihood of perceptual epenthesis follows exactly the presumed universal markedness between onset clusters. The universally most marked clusters (with a sonority fall) showed the highest likelihood of perceptual epenthesis, and the clusters that are universally least marked (with a sonority rise) were least likely to undergo perceptual epenthesis. Berent et al. conclude that this result shows that ‘English speakers are ... sensitive to the markedness of onset clusters that are unattested in their lexicon’, and hence that ‘speakers are equipped with universal preferences regarding the structure of language’ (p. 599). Their subjects had grammatical knowledge that they could not have acquired from their language exposure. (See Moreton 2002a, 2002b for more similar results.)

A question that needs answering is what the source of this knowledge is if it does not come from exposure. There are at least two possibilities. The first is that this reflects an innate property of linguistic competence – i.e. this knowledge is pre-programmed into the mind. The second option is that this is based on the general acoustic properties of the different sound sequences, and on the manner in which human auditory perception works – i.e. it is possible that the physical acoustic difference between [lbif] and [ləbif] is just smaller than that between [bdif] and [bədif], and that this is the source of the observed effect. Berent et al. specifically argue against the second possibility. They argue that, if the perceptual difference between [lbif] and [ləbif] is really smaller than that between [bdif] and [bədif], then their subjects should not only be more like to misclassify [lbif] as disyllabic, but also to misclassify [ləbif] as monosyllabic. This does not agree with their findings – their subjects were more likely to classify [ləbif] accurately as disyllabic, and more likely to misclassify [bədif] as monosyllabic. They interpret this as evidence that the greater amount of perceptual epenthesis in [lbif] is not due merely to a smaller acoustic difference between [lbif] and [ləbif] (Berent et al. 2007:612). They also show that, in an experimental design that encourages listeners to pay attention to the fine acoustic detail of a stimulus, subjects become quite good at accurately perceiving both [lbif] and [bdif], and that the difference between the cluster types disappear (Berent et al. 2007:620). It is only when the task requires of listeners to perform phonological processing that the difference between the cluster types emerges. Berent et al. interpret this as evidence that this difference is not rooted in phonetic similarity but in abstract phonological preferences.

For the purposes of this paper, it is not really important whether these preferences are innate or grounded in phonetic perceptibility. The important point is that the subjects in their experiment respond in ways that cannot be explained from the statistical patterns of the language

⁴ Kabak and Idsardi (2007) report similar results with Korean listeners. However, they found perceptual epenthesis only with some non-attested clusters, which shows that a more careful nuanced interpretation of the phenomenon of perceptual epenthesis is required.

to which they were exposed, and specifically in ways that reflect cross-linguistic typological generalizations. This could be interpreted as evidence of Universal Grammar influencing their response patterns, irrespective of whether Universal Grammar is taken to be pre-programmed and innate, or externally grounded in phonetic perceptibility (Hayes et al. 2004, Hume and Johnson 2001).

If grammar were simply the sum total of language experience, then successful grammar learning must be possible without assuming much if any pre-existing grammatical structure. Bybee (2006:711) says the following with regard to grammar learning in a strictly usage-based theory of language: '[T]he general cognitive capabilities of the human brain, which allow it to categorize and sort for identity, similarity, and difference, go to work on the language events a person encounters, categorizing and entering in memory these experiences. The result is a cognitive representation that can be called a grammar.' Grammar learning should thus be achievable by means of 'general cognitive capabilities' and should not require any pre-existing knowledge about linguistic structures. Results such as those of Berent et al. discussed above already show that this strict interpretation of usage-based grammar learning is problematic. There is also evidence from research that deals specifically with grammar learning that shows that successful grammar learning from exposure to language use is only achievable if a certain amount of specialized, pre-existing linguistic knowledge is allowed.

Typical learning algorithms in OT (Boersma and Hayes 2001, Tesar and Smolensky 1998) provide the learner with a pre-existing set of constraints, and the task of the learner is simply to determine the ranking between the constraints. Hayes and Wilson (to appear) propose a different phonological learning algorithm where the learner is also given the task of learning the constraints themselves from statistical properties of the data. The Hayes and Wilson approach is obviously more in line with the general assumptions of a strictly usage-based account of grammar learning. However, not even their algorithm can do without assuming pre-existing linguistic knowledge. Even for very simple phonotactic learning, their model needs to assume both a set of distinctive features and a general template of the kind of constraint that must be learned. When they apply their algorithm to learn grammars that go beyond mere phonotactics (like the non-local dependencies in harmony and stress placement), the algorithm fails to learn successfully if it is not augmented by even more non-learned knowledge such as the segmental tiers of autosegmental phonology and the grid structure of metrical phonology. It is therefore possible that constraints can be learned from language usage data. However, as Hayes and Wilson show, this cannot happen without pre-existing language specific cognitive structures.

Peperkamp et al. (2006) come to the same conclusion. They develop a purely statistical learning algorithm to learn allophonic rules. The algorithm looks for pairs of sounds that occur in complementary distribution, and it then learns a rule to relate the sounds in a pair to each other. They then create a corpus based on about 42,000 short child-directed French utterances. They marked up the corpus to reflect two allophonic processes (palatalization of /k, g/ before high front vowels, and devoicing of sonorants before voiceless obstruents). They found that their purely statistical learning algorithm is very successful at learning the two allophonic processes when presented with this corpus. However, the algorithm also learns many spurious allophonic rules such as the following: The semi-vowel [ɥ] occurs only as the last member of a syllable onset and hence before vowels (as in *plui* [plɥi] 'rain'). The vowel [œ] occurs only in closed syllables and hence before consonants (as in *peur* [pœʁ] 'fear'). These two sounds thus stand in structurally the same relation to each other as the voiced and voiceless counterparts of sonorants

– they are all in complementary distribution. A purely statistical learning algorithm cannot distinguish between linguistically relevant complementary distribution and spurious examples such as these. Their conclusion is that a successful grammar learning algorithm must incorporate ‘linguistic filters’. A general statistical algorithm can identify statistical dependencies in the data. But a pre-existing linguistic filter has to weed out the linguistically spurious from the relevant statistical generalizations.⁵ See also Moreton (2006) and Wilson (2006) for arguments that phonological learning needs access to more than frequency statistics.

This research shows that, even if grammar is the ‘cognitive organization of one’s experience with language’ (Bybee 2006:711), this experience is organized according to linguistic principles that cannot be extracted directly from this experience. There must be a core of linguistic knowledge that is independent from experience.

The last piece of evidence in defense of grammar as independent from usage frequency that I discuss here comes from a language processing experiment. This is particularly relevant to the current paper, as I later use similar processing experiments as evidence about the formal properties of grammar.

It has been illustrated many times that different kinds of usage frequency statistics influence tasks like word-likeness judgments. The frequency statistics that are used most often in the psycholinguistic literature are lexical neighborhood density (LND) and transitional probability (TP) (Bailey and Hahn 2001; Luce 1986; Luce and Pisoni 1998; Newman et al. 1997; Vitevitch and Luce 1998, 1999). The LND of a token is a measure that takes into consideration both the number of words in the lexicon that are similar to the token to some specified degree, and the usage frequency of these words. A token that is similar to many and/or frequent real words has a high LND, and a token similar to few and/or infrequent words has a lower LND. The TP of two sounds is a measure of the likelihood of those sounds occurring in a specific order. If the sequence [xy] occurs in more and/or more frequent words than the sequence [xz], then the TP of [xy] will be higher than that of [xz]. There is ample evidence from the literature that non-words with high values for their frequency statistics usually receive higher word-likeness ratings than non-words with lower values (Bailey and Hahn 1998, 2001; Coleman and Pierrehumbert 1997; Frisch et al. 2001; Pierrehumbert et al. 2004; Treiman et al. 2000).

Results like these could be interpreted as showing that at least when it comes to tasks like word-likeness rating, there might not be a need for grammar independent from usage-frequencies. However, although there is a correlation between these frequency statistics and word-likeness ratings, it is also the case that this correlation is far from perfect. There is usually a large amount of variation that is not explained by the frequency statistics (r^2 -values in these studies are often very low, sometimes even lower than .10). This creates the expectation that there are factors other than just these statistics that might be involved. One possibility is that grammar might also contribute to the observed effects. Frisch and Zawaydeh (2001) investigated this possibility. Arabic verbal morphology, like that of other Semitic languages, is based on tri-consonantal verbal roots. These verbal roots are subject to an OCP-type restriction that prohibits the occurrence of homorganic consonants in the same root – i.e. a root such as */f-m-s/ is not

⁵ These ‘linguistic filters’ will have to be formulated carefully so that they do not eliminate real but unexpected allophonic relations. Peperkamp et al. are conscious of this, which is why they argue for filters that are typologically motivated. The specific filters that they use in their learner compare the two sounds that are in complementary distribution, and reject pairs of sounds that are too different from each other.

tolerated because of the two labials (Frisch et al. 2004; McCarthy 1986, 1994; Pierrehumbert 1993). Frisch and Zawaydeh chose roots that do not correspond to actual roots of Arabic. Some of these roots contained homorganic consonants and others did not; in other words, some violated the OCP-Place constraint and others did not. They also calculated two kinds of frequency statistics for their tokens, namely ‘expected probability’ (similar to transitional probability) and neighborhood density. They conjugated these non-word roots as verbs, and presented them to Arabic speakers, with the task to rate each non-word for its word-likeness. They found a highly significant effect of OCP-violation on the ratings – non-words that violated the OCP were rated significantly less well than non-words that did not violate the OCP. However, crucially, they found no effect of either expected probability or neighborhood density on their results – ‘OCP-place violation influenced the ratings, but the manipulation of expected frequency and neighborhood density did not’ (2001:98). They hence conclude that ‘knowledge of phonotactic constraints is independent from simple lexical statistics’ (2001:104).

Berent, Everett and Shimron (2001), Berent, Shimron and Vaknin (2001), and Berent et al. (2002) come to a similar conclusion based on word-likeness rating experiments with Hebrew listeners. There is also similar evidence for an independent contribution of grammar to speech processing in other kinds of tasks. Moreton (2002), Pitt (1998), Berent et al. (2007), Dupoux et al. (2001) and Coetzee (2005) all show that grammar influences perception. Berent, Shimron and Vaknin (2001) and Berent et al. (2002) show in a similar manner that grammar influences reaction times in lexical decision. Zuraw (2007) shows that the way in which Tagalog speakers applies a morphological process to nonce words cannot be accounted only by frequency statistics.

Based on the results discussed in this section, I assume in the rest of this paper that formal grammar, though not completely independent from usage statistics, can also not be completely subsumed by usage statistics. I also assume that both usage frequency statistics and grammar can influence tasks like phoneme identification, word-likeness rating, and lexical decision. If we control for the possible influence of usage statistics in a responsible manner in experimental design, then the results of these kinds of psycholinguistic experiments can be interpreted as giving evidence about grammar per se.

It is well known from the literature, such as those discussed above, how frequency statistics influence tasks like phoneme identification, word-likeness judgments and lexical decision. In general, higher frequency statistics favor a token in phoneme identification and in word-likeness judgments. An acoustically ambiguous token is more likely to be perceived as a token with higher frequency statistics. Similarly, a token with higher frequency statistics will in general receive higher word-likeness ratings than a token with lower frequency statistics. In lexical decision, however, higher frequency statistics count against a token – i.e. a non-words with higher frequency statistics will be identified as a non-word more slowly than a non-word with lower frequency statistics. In this paper, I assume that grammar can influence these tasks in a similar manner. Specifically, with regard to phoneme identification, I assume that a token that is acoustically ambiguous between a more and a less well-formed percept, is more likely to be perceived as the more well-formed token (Moreton 2002). With regard to word-likeness ratings, I assume that a non-word that is grammatically more well-formed will be rated as more word-like than a form that is grammatically less well-formed (Frisch and Zawaydeh 2001). Finally, with regard to lexical decision: Since I expect a more well-formed non-word to be rated as more word-like, I also expect that listeners will consider a more well-formed token more seriously as a

possible word and will hence be slower to reject it as a non-word (Berent, Shimron and Vaknin 2001). I do not formulate a detailed model of how grammar impacts these processing tasks, of where in the time-course of processing grammar is active, or of how grammar and frequency statistics interact during processing. The goal of this paper is to show that grammar in addition to frequency statistics influences processing. The development of a comprehensive processing model that incorporates both frequency statistics and grammar, is left for future research.

4. RESTRICTIONS ON WORDS OF THE FORM [sC_iVC_i] IN ENGLISH

In this section, in order to illustrate the two predictions about grammar discussed above in section 2, I discuss a restriction on words of the form [sC_iVC_i] in English, and develop a standard OT account of this restriction. I first present the facts about [sC_iVC_i] forms in English, and develop a basic OT grammar to account for these facts. I then motivate the constraints used in the OT account, and finally consider how the grammar can be learned from frequency data.

4.1. THE BASIC [sC_iVC_i]-GRAMMAR OF ENGLISH

The English lexicon contains words of the form [sC_iVC_i] where both C's are [t], but not where both C's are [k] or [p] – i.e. *state* is a word but *skake* and *spape* are not. A question that needs to be answered is whether the absence of words like *skake* and *spape* is accidental, or whether forms like these are ungrammatical.⁶ The standard assumption in the literature is that forms like these are excluded on grammatical grounds (Browne 1981; Clements and Keyser 1983; Davis 1982, 1988a, 1988b, 1989, 1991; Fudge 1969; Lamontagne 1993: Chapter 6), and I follow the tradition in the literature.^{7, 8} The restriction against these forms is part of a larger set of

⁶ The clearest way to show that a form is ungrammatical is to show that it is actively avoided or repaired. For instance, in English a tautosyllabic syllabic [voiceless obstruent + voiced obstruent] sequence is ungrammatical, as evidenced by the differing realizations of the plural – *backs* [bæks] vs. *bags* [bægz]. But what about a word beginning with the sequence [bn-]? Since no such lexical entry exists and since this sequence cannot be created by any morphological process of English, there is no active process that repairs such sequences on a par with the obstruent voicing assimilation observed in the plural. Even so, forms like [bnik] are standardly assumed to be ungrammatical in English (Halle 1962). What is required is a way to distinguish between accidental lexical gaps and principled lexical gaps. It is not easy to formalize a principle that can be used for this purpose. However, as a first approximation I suggest the following: Lexical gaps that can be stated in terms general phonological properties should be treated as principled. [bnik] can be ruled out since it belongs to a class of forms with an onset sonority rise that is not steep enough – including forms like [knik, pnik, gmik, ...]. However, the non-word [blik] should be considered an accidental gap. It is not a member of any general class of forms that are absent from the lexicon. Since [spVp] and [skVk] form part of a larger, formally definable class of forms that are absent from the lexicon (Fudge 1969), these forms should also be considered principled gaps. (See Frisch and Zawaydeh 2001 for a similar distinction between accidental and principled lexical gaps.)

⁷ This same restriction is also active in German. Twaddell (1939, 1940) compiled a list of all the mono- and bi-syllabic words in Duden's 1936 German dictionary. This resulted in a list of approximately 37,500 tokens. He then calculated the frequency of all [(C)(C)(C)V(C)(C)(C)]-sequences (where [V] is a stressed vowel). He found 51 words with the sequence [stVt], but none with the sequence [skVk] or [spVp].

⁸ Although the English lexicon does not include words of the form [spVp] or [skVk], there are acronyms of this form. For instance, SPEP is the acronym for the 'Society for Phenomenology and Existential Philosophy' and SCAC for the 'Southern Collegiate Athletics Conference'. It is hard to decide whether the existence of

restrictions on the consonants that can co-occur in the onset and coda of syllables (Fudge 1969). For instance: (i) A mono-syllabic word cannot begin with [Cr] and end in [r] – i.e. **frer*, **krer*. (ii) A mono-syllabic word cannot begin in [s] plus a nasal and end in nasal – i.e. **snam*, **smang*. (iii) A mono-syllabic word cannot begin in [sl] and end in [l] – i.e. **slul*, **slel*, etc. I focus here only on the restriction as it applies to voiceless stops in the configuration [sCVC].

I assume a markedness constraint against each of these three kinds of forms, i.e. *stVt, *skVk, and *spVp, and motivate the existence of these constraints below. But first I consider how these constraints are ranked in English. Since English tolerates words of the form [stVt], we know that *stVt ranks below all relevant faithfulness constraints. On the other hand, since English does not tolerate words such as [skVk] or [spVp], we know that both *skVk and *spVp rank higher than some relevant faithfulness constraint. This gives evidence for the ranking { *spVp, *skVk } >> Faithfulness >> *stVt. But can we rank *spVp and *skVk relative to each other? There is ample evidence from elsewhere in English phonology that English tolerates the co-occurrence of dorsals better than that of labials. For instance, English allows words like [skVg] but not like [spVb] – cf. *skag* vs. **spab*. Similarly, English allows words of the form [skVXk] where [X] is a nasal or liquid, but not of the form [spVXp] – cf. *skunk*, *skulk* vs. **spump*, **spulp*. This suggests that *spVp in all likelihood outranks *skVk. The full mini-grammar of English is then *spVp >> *skVk >> Faithfulness >> *stVt. The tableau in (1) shows that this ranking correctly reflects the fact that [stVt], but not [skVk] or [spVp], is a possible word of English. Since English has no alternations to resolve violations of *sCVC-constraints, we do not know what the relevant faithfulness constraint in this tableau should be. It can be any faithfulness constraint that can be violated to avoid violation of *skVk and *spVp.

- (1) [stVt] is a possible word, but [skVk] and [spVp] are not

			*spVp	*skVk	IDENT[place]	*stVt
/stVt/	a.	☞ [stVt]				*
	b.	[stvk]			*!	
/skVk/	c.	[skVk]		*!		
	d.	☞ [skvt]			*	
/spVp/	e.	[spVp]	*!			
	f.	☞ [spvt]			*	

This OT account presents examples of both predictions discussed in section 2. First, although [stVt] is correctly predicted as a possible word, [stVt] is also not perfect – it violates *stVt. This presents a problem: since there is no alternation in English that involves *stVt, we do not have direct evidence that *stVt is part of English grammar. We know that [stVt]-forms are allowed in English, but this can be interpreted in two ways: either *stVt ranks lower than

acronyms like these imply that [spVp] and [skVk] are possible words of English. Acronyms function like proper names, and proper names are often exempt from constraints that apply to the rest of the lexicon. It is also true that, although acronyms like these do exist, they are rare and unlikely to be included in the mental lexicon of the average speaker of American English. I conducted a small study to determine whether the lexicon of the average American English speaker contains acronyms like these. Seventeen native speakers of American English were recruited from the University of Michigan. I explained to the participants what an acronym is, and presented them with eight real acronyms of the form SPVP, SKVK, SCVC, SCVK and SKVC. I then asked them to write down any such acronyms that they know. Not one of the participants knew any such an acronym.

faithfulness in English (the standard OT assumption), or *stVt is not part of English grammar at all.

This kind of situation where even grammatical forms violate at least some markedness constraints is the rule rather than the exception in OT. This is a direct result of the standard OT assumption that the constraint set is universal. Even markedness constraints that are violated freely in some language are part of the grammar of that language – they are simply ranked very low. However, this assumption is not a necessity. If we were to remove *stVt from the tableau in (1), the grammar would still make the correct predictions about English – [stVt] would still be a possible word, and [skVk] and [spVp] would still be impossible. It is therefore also possible that constraints that are completely ‘inactive’ in some language are dropped all together from the constraint hierarchy of that language. This is an important question, not only because it involves one of the central assumptions of OT, but also because it speaks to the core of what the grammar of an individual language is like. Do individual languages rid themselves of the inactive parts of Universal Grammar, or are the inactive parts still preserved in some form in the grammar of individual languages?

The data traditionally used in phonology cannot help us to decide between these two possibilities. These data take on the form ‘/stVt/ → [stVt] is a grammatical mapping in English’. But traditional data cannot distinguish between the statements ‘/stVt/ → [stVt] is grammatical because it does not violate any constraint of English’ and ‘/stVt/ → [stVt] is grammatical even though it violates *stVt’. In section 5, I show how we can use the results of a phoneme identification experiment to distinguish between these two possibilities.

The OT account above also provides an example of the prediction that there are degrees of ungrammaticality. Since both *spVp and *skVk rank higher than faithfulness, it is correctly predicted that neither [spVp] nor [skVk] should be allowed as words of English. However, the ranking *spVp >> *skVk implies that [spVp] is somehow less well-formed than [skVk]. This kind of situation is again very common in OT analyses. It is often the case that a ranking is implied between markedness constraints that are never violated by some language; that is the constraints that rule out impossible words as ungrammatical.

The ranking between *spVp and *skVk is not necessary to account for the surface sound patterns of English – even if the ranking is removed, [spVp] and [skVk] are still predicted as impossible words. This also involves core questions about the structure of the grammar of individual languages. Is the grammar of some language only as complex as necessary to distinguish grammatical and ungrammatical forms from each other? Or does the grammar of an individual language contain more structure than this?

Once again this question cannot be answered by relying on the data traditionally used in phonology. These data simply tell us that neither [spVp] forms nor [skVk] forms are attested as words in English. But they are quiet about any well-formedness relationship between these two forms. In sections 6 and 7, I discuss the results of a word-likeness judgment experiment and lexical decision experiment that show that English speakers do treat [spVp]-forms as less well-formed than [skVk]-forms.

4.2. MOTIVATING THE *sCVC-CONSTRAINTS

The *sCVC-constraints can be viewed as the local conjunction between OCP-type constraints against multiple occurrences of the same segment, and a constraint against the sequence [s+stop] (Coetzee 2004:402-420; see Baertsch and Davis 2003 for an alternative).

Languages often show restrictions on the co-occurrence of identical (or highly similar) consonants within some local domain. This has been demonstrated for languages as diverse as Arabic (Frisch et al. 2004; McCarthy 1994; Pierrehumbert 1993), Japanese (Kawahara et al. 2006), Russian (Padget 1995), Muna (Coetzee and Pater to appear), Javanese (Mester 1986; Yip 1989; Uhlenbeck 1949), Rotuman (McCarthy 2003), French, Latin and English (Berkley 2000). Although the details of the restrictions differ between the languages, it is clear that there has to be a general family of constraints that militate against the occurrence of identical or highly similar consonants in some local context. Alderete (1997) and Itô and Mester (1998) analyze restrictions like these using local self-conjunction (Smolensky 1995) of markedness constraints. However, Coetzee and Pater (to appear) assume that OCP-type constraints are primitive constraints that are not built up by self-conjunction. I follow the primitive constraint option here, however for the purposes of this paper it would not make any difference if the local self-conjunction approach were taken.

(2) OCP-type constraints for English⁹

*_vt: Do not allow two [t]'s separated by only a vowel.

*_vk: Do not allow two [k]'s separated by only a vowel.

*_vp: Do not allow two [p]'s separated by only a vowel.

These constraints are not stated in terms syllable structure, since English lacks not only monosyllabic words of the form [skV_vk] and [spV_vp], but there are in fact no words in English with these sequences even with an intervening syllable boundary – i.e. [spV.pV] is completely unattested, and with the exception of a few proper names (such as *Skakel*), so too is [skV.kV].¹⁰

There are two views in the literature about why [s+stop]-sequences are marked. Under one view, these sequences are considered true consonant clusters (Gouskova 2001; Hayes 1985:148-149; Hockett 1955:152-153; Kahn 1980:42-43). If these sequences are consonant clusters, they violate the Sonority Sequencing Principle (SSP), since the sonority then falls from the fricative [s] to the stop. Under the alternative view, [s+stop]-sequences are not consonant clusters, but rather complex segments (Broselow 1991; Firth 1935, 1935-1937:543; Fudge 1969; Fujimura and Lovins 1978:112; Lamontagne 1993: Chapter 6; Selkirk 1982; Spang-Hanssen 1959:158-161). If these sequences are complex segments, then they avoid violation of the SSP (as single segments they have only one sonority value). However, if they are complex segments then they are marked as such. Irrespective of how these sequences are viewed, it is clear that they

⁹ These constraints can be stated more generally as ‘do not allow two coronals separated by only a vowel’, etc. With a few exceptions, these broader definitions of *skV_vk and *spV_vp are also unviolated in English (cf. *scag*, *skeg*, *spume*, *spam*). Because of these exceptions, I have opted for the more restrictive definition in the text.

¹⁰ A second motivation for this comes from Twaddell’s German data. As explained in footnote 7, Twaddell did not take syllable boundaries into account. He therefore found no examples of German words with the sequences [skV.kV] or [spV.pV].

are marked. I do not take a stance on which of these approaches is correct, but just assume that [s+stop]-sequences violate some markedness constraint.

- (3) * $[s+stop]_{\sigma}$ Do not allow the sequence [s+stop] in a syllable.

English tolerates violation of each of the constraints in (2) and (3) individually – *toot* (* t_vt), *cake* (* k_vk), *pope* (* p_vp), *sky* (* $[s+stop]_{\sigma}$), *whisk* (* $[s+stop]_{\sigma}$), etc. However, what English does not tolerate is the violation of some combinations of these constraints – English does not tolerate violation of * $[s+stop]_{\sigma}$ together with either * k_vk or * p_vp . Based on this, I propose that the *sCVC-constraints are just the local conjunction (Smolensky 1995) of the constraints in (2) with * $[s+stop]$.¹¹ (See Itô and Mester 1997 who analyze final devoicing in German in a similar manner.)

- (4) * $stVt$: Do not violate * $[s+stop]_{\sigma}$ and * t_vt in the same syllable.
 * $skVk$: Do not violate * $[s+stop]_{\sigma}$ and * k_vk in the same syllable.
 * $spVp$: Do not violate * $[s+stop]_{\sigma}$ and * p_vp in the same syllable.

4.3. LEARNING THE GRAMMAR

I argued above for the ranking * $spVp$ >> * $skVk$ >> * $stVt$. Since [stVt] words exist while [spVp] and [skVk] words do not, it is easy to imagine how English speakers could arrive at the ranking { * $spVp$, * $skVk$ } >> * $stVt$. But how can they learn the ranking between * $spVp$ and * $skVk$ in the absence of specific evidence about the ranking between these two constraints?

I show here how a ranking between markedness constraints can be learned even if these constraints are freely violated in the learning data. The ranking that is learned reflects the frequency with which the constraints are violated, such that constraints that are violated less often rank higher. In spite of the fact that * p_vp , * k_vk , and * t_vt are freely violated in English, learners can use information on how frequently these constraints are violated to learn a ranking between them. CELEX has a special sub-corpus that lists the frequency per million of syllables. I calculated in this sub-corpus the frequency of syllables that contain two [t]'s, [k]'s or [p]'s separated by only a vowel. $_{\sigma}[\dots tVt\dots]_{\sigma}$ has a frequency of 5348 per million, $_{\sigma}[\dots kVk\dots]_{\sigma}$ of 695, and $_{\sigma}[\dots pVp\dots]_{\sigma}$ of 235. Forms with two [t]'s are hence most frequent, then forms with two [k]'s, and then forms with two [p]'s. See also Berkley (2000) for similar statistics about the English lexicon. Using frequency data such as these, learners can acquire the ranking * p_vp >>

¹¹ A word like *cas*k [kæsk] does violate [s+stop]_σ but not * k_vk – since the two [k]'s are separated by more than just a single vowel from each other. Consequently, [kæsk] also does not violate * $skVk$. [skæk], on the other hand, violates both [s+stop]_σ and * k_vk , and consequently also * $skVk$. The analysis presented here therefore does not incorrectly rule out *cas*k as ungrammatical. On the other hand, this analysis also does not rule out as ungrammatical a form like **pas*p. What is needed are less strict versions of * p_vp , * k_vk , and * t_vt that are violated when there are two segments separating the identical stops from each other – i.e. something like * $p_{vc}p$, * $k_{vc}k$, and * $t_{vc}t$. The local conjunction of these constraints with * $[s+stop]_{\sigma}$ will be violated by forms like *test*, *cas*k, and **pas*p. The conjoined constraint [* $[s+stop]$ & * $p_{vc}p$] will then rank above some relevant faithfulness constraint too rule out **pas*p as a possible word. On the other hand, [* $[s+stop]$ & * $k_{vc}k$] and [* $[s+stop]$ & * $t_{vc}t$] will rank below all relevant faithfulness constraints so that *test* and *cas*k are allowed. See Berkley 2000 for evidence that [kVk], [pVp], and [tVt] are more restricted than [kVck], [pVCp], and [tVCt].

*k_vk >> *t_vt. Once this ranking has been established, the principle of ‘ranking preservation’ (Itô and Mester 1998) dictates that all constraints formed by the local conjunction of these constraints with some other constraint preserve this ranking. When *spVp and *skVk are formed by the local conjunction of *[s+stop] with *p_vp and *k_vk, respectively, the ranking *spVp >> *skVk is hence inherited from the ranking *p_vp >> *k_vk. The rest of this section fleshes out this proposal.

Pater (2005, also Pater and Coetzee 2005) proposes an extension to the Biased Error-Driven Constraint Demotion (BECD) algorithm of Tesar and Smolensky (1998; also Prince and Tesar 2004) that gives the algorithm the power to learn frequency based rankings between markedness constraints. In the BECD, every time the learner encounters a new form (hears a new word), this word is sent through the learner’s current grammar. If the current grammar generates a form that is different from the learning datum (i.e., an error is generated), a Winner~Loser pair is created and added to the Support (Tesar 1998; Prince and Tesar 2004). The Support is a set of these Winner~Loser pairs from which the learner learns the grammar. In the Support, constraints that prefer the Winner (the intended output) are marked with a W, and constraints that prefer the Loser (the candidate incorrectly generated by the learner’s current grammar) are marked with an L. Once the Support has been created, the algorithm, given in (5) goes into effect.

- (5) Biased Error-Driven Constraint Demotion
 - i. Find all the constraints that prefer no Losers (i.e. with only W’s in their columns).
 - ii. If any of these are markedness constraints, install them in the current stratum, and return to step i.
 - iii. If there are no available markedness constraints, install faithfulness constraints that prefer winners, and return to step i.
 - iv. If there are no faithfulness constraints that prefer winners, install faithfulness constraints that prefer no Losers, and return to step i.

Pater’s innovation is to suggest that learners are lexically conservative during the process of creating the Support. Every time the learner’s grammar generates an error so that a Winner~Loser pair is created, the faithfulness constraints that favor the Winner are indexed only to the lexical item concerned. During this stage of learning, learners therefore assume that all violations of markedness are exceptional or specific to the lexical item at hand.

To see how this works, let us consider a small part of the English lexicon that preserves the basic relative frequencies mentioned above – see (6). Assuming the initial state ranking Markedness >> Faithfulness (Smolensky, 1996; Itô and Mester, 1999; Tesar and Smolensky, 2000; Hayes, 2004; Prince and Tesar, 2004), all of these words will initially result in an error and hence in the creation of a Winner~Loser pair. The Support that will be created from this mini lexicon is given in (7). For the sake of concreteness, I assume that the relevant faithfulness constraint is IDENT[place]. However, in the absence of evidence from alternations, it could be any faithfulness constraint that can be violated to avoid the markedness violations.

- (6) Mini lexicon

<i>toot</i>	<i>cook</i>	<i>pope</i>
<i>tote</i>	<i>cake</i>	
<i>teat</i>		

(7) Support

Winner~Loser	* _t v _t	* _k v _k	* _p v _p	ID _{toot}	ID _{tote}	ID _{teat}	ID _{cook}	ID _{cake}	ID _{pope}
toot~took	L			W					
tote~toke	L				W				
teat~teek	L					W			
cook~coot		L					W		
cake~cate		L						W	
pope~pote			L						W

Once this Support has been created, the BECD algorithm goes into action. The first step is to collect all the constraints that prefer no Losers. In the Support in (7), this is the full set of lexically indexed faithfulness constraints. Next, the algorithm needs to determine which of these constraints to rank first. Since the set contains no markedness constraints, the second step in the algorithm does not apply, and we move to the third step. At this point, the algorithm has six constraints that it can rank. Prince and Tesar (2004:257) propose that the choice between faithfulness constraints is made by identifying the smallest set of faithfulness constraints whose ranking will ‘free up’ a markedness constraint for ranking.¹² They refer to this set as the ‘smallest effective F set’, defined in (8).

(8) Smallest Effective F set

When placing faithfulness constraints in a hierarchy, first place the smallest set of faithfulness constraints that free up some markedness constraint.

In the Support in (7) above, the smallest effective F set is {IDENT_{pope}}. Once it has been ranked high, the learner’s grammar no longer generates an error when presented with the token *pope* so that the L from the column for *_pv_p will disappear from the Support. On the next round of learning, *_pv_p therefore becomes available for ranking. Two faithfulness constraints have to be ranked to free up *_kv_k, and three to free up *_tv_t. Based on this, the learner hence installs IDENT_{pope} in the highest stratum, and removes the last line from the Support – since his/her grammar no longer generates an error when presented with *pope*. The grammar learned thus far and the updated support are given in (9).

¹² Ranking a faithfulness constraint ‘frees up’ a markedness constraint if the ranking results in all L-marks for the markedness constraint being removed from the Support. On the next round of learning, the markedness constraint therefore becomes available for ranking.

(9) Grammar and support thus far

Grammar: $IDENT_{pope} \gg \dots$

Support

Winner~Loser	* _t v _t	* _k v _k	* _p v _p	ID _{toot}	ID _{tote}	ID _{teat}	ID _{cook}	ID _{cake}
toot~took	L			W				
tote~toke	L				W			
teat~teek	L					W		
cook~coot		L					W	
cake~cate		L						W

The BECD algorithm now applies again, first collecting the set of constraints that prefer no Losers. This time, this set includes the markedness constraint *_pv_p. In the second step of the algorithm, this constraint is hence installed on the next stratum and removed from the Support, as in (10).

(10) Grammar and support thus far

Grammar: $IDENT_{pope} \gg *p_v p \gg \dots$

Support

Winner~Loser	* _t v _t	* _k v _k	ID _{toot}	ID _{tote}	ID _{teat}	ID _{cook}	ID _{cake}
toot~took	L		W				
tote~toke	L			W			
teat~teek	L				W		
cook~coot		L				W	
cake~cate		L					W

Next, the BECD again collects the constraints that prefer no Losers. This time, this set again contains only the faithfulness constraints, and the smallest effective F set needs to be found. To free up *_kv_k, two faithfulness constraints need to be ranked, while three are necessary to free up *_tv_t. The two that will free up *_kv_k is hence installed, and the Support is trimmed down accordingly, as in (11).

(11) Grammar and support thus far

Grammar: $IDENT_{pope} \gg *p_v p \gg \{ IDENT_{cook}, IDENT_{cake} \} \gg \dots$

Support

Winner~Loser	* _t v _t	* _k v _k	ID _{toot}	ID _{tote}	ID _{teat}
toot~took	L		W		
tote~toke	L			W	
teet~teek	L				W

At this point it should be clear that *_kv_k will be ranked next, followed by the last three faithfulness constraints, and finally by *_tv_t. The final grammar that is learned is given in (12).

For more on the role of the indexed faithfulness constraints in this grammar, see Pater (2005). All that is relevant for the purposes of this paper, is the ranking between the three markedness constraints – $*p_{vp} \gg *k_{vk} \gg *t_{vt}$. This ranking was learned although all three constraints were freely violated in the learning data, and the ranking between the constraints reflect the frequency with which the constraints are violated in the learning data, with constraints violated less frequently ranked higher. This is achieved by a conspiracy between Pater’s proposal that faithfulness constraints are treated as word-specific during learning and Prince and Tesar’s idea about the smallest effective F set.

(12) Final grammar

$$ID_{pope} \gg *p_{vp} \gg \{ID_{cook}, ID_{cake}\} \gg *k_{vk} \gg \{ID_{toot}, ID_{tote}, ID_{teat}\} \gg *t_{vt}$$

Based on frequencies in the English lexicon, English learners therefore acquire the ranking $*p_{vp} \gg *k_{vk} \gg *t_{vt}$. I argued above that the $*sC_iVC_i$ constraints are the local conjunction between $*[s+stop]$ with each of $*p_{vp}$, $*k_{vk}$ and $*t_{vt}$. Itô and Mester (1998) propose a limit on the ranking between locally conjoined constraints that they term ‘ranking preservation’, defined in (13). This principle preserves the ranking between conjuncts of conjoined constraints in the conjoined constraints. That is, since $*p_{vp} \gg *k_{vk} \gg *t_{vt}$, all constraints formed by the conjunction of these constraints with some other constraint must preserve the ranking between these constraints. From this, the ranking assumed above between the three $*sC_iVC_i$ constraints follows directly, as shown in (14).

(13) Ranking preservation

Let LC_1 and LC_2 be two locally conjoined constraints with C_1 one of the conjuncts of LC_1 and C_2 one of the conjuncts of LC_2 . If $C_1 \gg C_2$, then $LC_1 \gg LC_2$.

(14) Ranking preservation applied to $*sC_iVC_i$ constraints

Learned ranking: $*p_{vp} \gg *k_{vk} \gg *t_{vt}$

Inferred ranking: $[*p_{vp} \& *s+stop] \gg [*k_{vk} \& *s+stop] \gg [*t_{vt} \& *s+stop]$
(*spVp) (*skVk) (*stVt)

Although the learning data that an English learner is exposed to contains no direct evidence for a ranking between $*spVp$ and $*skVk$, the learner generalizes the ranking that he/she has learned from frequency data elsewhere to the ranking between these two constraints. This is in agreement with Frisch and Zawaydeh’s (2001:104-105) view of the relation between frequency and grammar. In their view grammar is ‘emergent’ and phonotactic constraints are learned ‘based on lexical patterns’. However, they also conclude that ‘knowledge of phonotactic constraints is independent from simple lexical statistics’. Learners rely on frequency information included in the learning data to acquire grammar. However, once the grammar has been acquired, it assumes a certain level of independence from this frequency information. This suggests a compromise between the usage-based and formal grammar approaches to language. As shown convincingly by a long line of research on usage-based theories of language, frequency information does play a role in language acquisition, use and processing. Here I assume that learners use such information in the initial stage of grammar learning. However, once they have learned the grammar based on this information, grammar can take on a life independent from this information.

The next three sections contain discussion of three experiments that I conducted to investigate the processing of $[sC_iVC_i]$ non-words in English. The first experiment is a phoneme

identification experiment, and shows that although [stVt] is a possible word of English there is a perceptual bias against it when it is compared to non-words of the form [stVk] and [stVp]. This is used to show that [stVt], although a possible word, does violate some constraint. The next two experiments are a word-likeness and a lexical decision experiment, and they are used to show that although neither [spVp] nor [skVk] is a possible word of English, [spVp] is judged as less acceptable than [skVk].

5. THE BEST IS NOT NECESSARILY PERFECT: PHONEME IDENTIFICATION

Several studies over the past few decades have argued that grammar can cause perceptual biases. Massaro and Cohen (1983), for instance, created two continua, [tli]~[tɹi] and [sli]~[sɹi], and presented them to listeners with the task of identifying the second sound. They found a bias toward [ɹ] on the [t]-continuum and towards [l] on the [s]-continuum. They interpret this as a perceptual bias caused by the phonotactics of English – English does not tolerate the sequences [tl-] and [sɹ-] word-initially.¹³ For similar evidence that grammar can bias perception, see Berent et al. (2007), Coetzee (2005), Dupoux et al. (1999), Dupoux et al. (2001), Kabak and Idsardi (2007), and Moreton (2002). When listeners are presented with a token that is ambiguous between two percepts, they are biased towards identifying the token as the more well formed of the two percepts. I conducted a phoneme identification experiment to test whether the restrictions against [sC_iVC_i]-words in English also cause biases like these. The results of this experiment show a perceptual bias against [stVt].

5.1. EXPERIMENTAL DESIGN

I constructed three sets of continua: (i) K~P: [skap]~[skak] and [spap]~[spak]; (ii) T~K: [stæk]~[stet] and [skæk]~[skæt]; (iii) T~P: [stap]~[stap] and [spap]~[spap]. The tokens on these continua were presented to listeners, whose task was to identify the final stop consonant. The expected perceptual biases in each condition are shown in Table 1.

¹³ It is, of course, possible to interpret this as the result of a frequency bias – English has more [tɹ-] and [sɹ-] initial words than [sɹ-] and [tl-] initial words. In order to evaluate the success of a purely frequency based explanation for Massaro and Cohen’s results, Pitt (1998) conducted a very similar experiment. In his experiment, he presented listeners with tokens from two continua, [bli]~[bli] and [gli]~[gli]. There is no frequency difference between [bɹ-] and [bl-] initial words in English, while [gɹ-] initial words are more frequent than [gl-] initial words. If frequency was responsible for the bias in Massaro and Cohen’s results, then the participants in Pitt’s experiment should respond [ɹ] more often on the [gli]~[gli] continuum than on the [bli]~[bli] continuum. Pitt found no difference in the response patterns on these continua, and therefore concludes that frequency cannot explain the response biases.

Condition	Continuum	Expected bias	Comment
K~P	[spap]~[spak]	Against [p]	Because [spap] violates *spVp
	[skap]~[skak]	Against [k]	Because [skak] violates *skVk
T~K	[skək]~[skɛt]	Against [k]	Because [skək] violates *skVk
	[stək]~[stɛt]	?	Toward [k] if [stɛt] violates *stVt
T~P	[spʌp]~[spʌt]	Against [p]	Because [spʌp] violates *spVp
	[stʌp]~[stʌt]	?	Toward [p] if [stʌt] violates *stVt

TABLE 1. Expected perceptual biases in the phoneme identification experiment

I discuss the K~P-condition to explain how perceptual bias was measured, but the other conditions were treated in the same way. The total percent of [p]-responses on the [skap]~[skak]-continuum and on the [spap]~[spak]-continuum was calculated. Since a bias towards [p] is expected on the [skap]~[skak]-continuum and a bias against [p] on the [spap]~[spak]-continuum, a higher total percentage of [p]-responses is expected on the [skap]~[skak]-continuum.

It is now possible to show how this experiment can be used to test whether [stVt]-tokens do or do not violate *stVt. In the K~P-condition, both *skVk and *spVp are active – there is a bias-causing constraint active on each continuum. In the T~K-condition, we know that *skVk is active on one of the continua. But we are not sure whether *stVt is active on the other continuum. If *stVt is not part of the grammar of English, then the bias-causing constraints in the K~P-condition (*spVp, *skVk) form a proper superset of the bias-causing constraints in the T~K-condition (*skVk). We would then expect a smaller bias-effect in the T~K-condition than in the K~P-condition. However, if *stVt is indeed part of the grammar of English, then there are two bias-causing constraints in the K~P-condition (*spVp, *skVk) and in the T~K-condition (*skVk, *stVt), and we would not necessarily expect a larger bias-effect in the K~P-condition. The T~P-condition can be interpreted in the same way as the T~K-condition.

5.2. TOKEN SELECTION

Lexicality causes perceptual biases in phoneme identification – when presented with a token ambiguous between a word and a non-word, listeners are biased towards perceiving the word (Ganong 1980). For this reason, tokens were selected such that the end-points of all continua were non-words. Frequency statistics, such as lexical neighborhood density (LND) and transitional probabilities (TP) (see the discussion section 3 above), also influence phonological processing. When presented with a token ambiguous between two percepts, listeners are biased towards perceiving the token with the higher values in terms of the frequency statistics (Connie et al. 1993; Newman et al. 1997). Since most research has focused on TP and LND, I calculated

the LND and TP of the endpoints in each of the six continua.¹⁴ All endpoints had TP's that were of the same order of magnitude. It was not possible to equalize the endpoint tokens also for LND. However, Vitevitch and Luce (1998, 1999) have shown that LND is less likely to influence tasks, such as phoneme identification, that do not require lexical access. All of the tokens were also monosyllabic, and it has been shown that lexical effects tend to decrease or even disappear completely in experimental setups where stimulus lists contain only monosyllabic tokens (Cutler et al. 1987). See also later in this section for more discussion on the possible role of frequency statistics in this experiment. A list of the tokens and all their frequency statistics is included in the appendix.

5.3. CONSTRUCTION OF STIMULI

I discuss the P~K condition as an example here, but the stimuli for the other conditions were created in the same manner. I recorded a phonetically trained native male speaker of American English reading the non-words [skap] and [spak] in the carrier sentence 'John said _____ again to me.' Four repetitions of each non-word were recorded in this way. The non-words [skap] and [spak] were then spliced from the carrier sentence, equalized for intensity, and stored in digital form. In each of the non-word tokens, the beginning and end of the syllabic nucleus were marked. The nuclei were then divided into ten intervals of equal duration, and the frequency and bandwidth of the first six formants were taken at the mid-point of each interval. The values obtained from the four [skap]-tokens were averaged and saved (the *skap*-file), and similarly for [spak] (the *spak*-file). The values obtained in this manner were used to create two twenty step continua – a continuum from [spak] to [spap], and a continuum from [skak] to [skap].

THE NUCLEUS. For the [spak] to [spap] continuum, the [spak]-endpoint token was synthesized as follows: The first 40% of the nucleus was synthesized using the values for the first four intervals from the *spak*-file, and the last 60% using the values for the last six intervals from the *spak*-file. The first 40% of the nucleus for the [spap]-endpoint was also synthesized with the values for the first four intervals from the *spak*-file. However, the last 60% of the

¹⁴ Frequency statistics were calculated using the CELEX database (Baayen et al. 1995). This database was 'Americanized' before the calculations were done. The 'Americanization' consisted primarily in changes in vowel quality and in the presence of post-vocalic /r/.

LND was calculated according to the method used by *inter alia* Vitevitch and Luce (1998, 1999) and Newman et al. (1997). The neighbors of a token are defined as any word that can be formed from the token by substitution, addition or deletion of one phoneme from the token. LND is calculated as follows: (i) Find all the neighbors for a token. (ii) Sum the log frequencies of all the neighbors. The LND therefore takes into account both the number of neighbors and their frequencies.

TP's were also calculated using a method very similar to that of Vitevitch and Luce (1998, 1999) and Newman et al. (1997). Consider the token [skek] as an example. For the sequence [ek] we can calculate the probability of an [e] being followed by [k], and the probability of a [k] being preceded by an [e]. To calculate the probability of [e] being followed by [k]: (i) Find all lemmas that contain an [e], and multiply the log frequency of each lemma with the number of [e]'s in the lemma. Then add up all of these multiplied log frequencies. (ii) Find all lemmas that contain the sequence [ek], and multiply the log frequency of each lemma with the number of [ek]-sequences in the lemma. Then add up all of these multiplied log frequencies. (iii) Finally, divide the summed log frequency of [ek]-sequences by the summed log frequency of [e]. The probability of a [k] being preceded by a [e] can be calculated in a similar manner.

nucleus for this endpoint was synthesized with the values for the last six intervals from the *skap*-file. The nuclei of the endpoints were therefore identical for the first 40% with unambiguous transitions from [p] into [ɑ]. They differed in their last 60%, having unambiguous transitions from [ɑ] into [k] and [p] respectively.

Between these two endpoint tokens, 18 intermediate tokens were synthesized. The first 40% of the nuclei in all 18 intermediate tokens was also synthesized with the values for the first four intervals from the *spak*-file. The first 40% of the nuclei for the intermediate and endpoint tokens was therefore identical. The last 60% of the nuclei of the 18 intermediate tokens was synthesized by mixing the values for the last six intervals from the *spak*-file and *skap*-file proportionally. This resulted in a continuum ranging from a token in which the final 60% of the nucleus agreed mostly with the [spak]-endpoint to a token in which it agreed mostly with the [spap]-endpoint. Putting together the endpoint tokens and the 18 intermediate tokens then resulted in a 20-step nucleus continuum between the endpoints [spak] and [spap].

Exactly the same procedure was used to construct a 20-step continuum from [skak] and [skap]. The only difference was that the first 40% of the nucleus for this continuum was synthesized using the values for the first four segments from the *skap*-file rather than the *spak*-file. The two nucleus continua were therefore identical in their last 60%, but differed in their first 40%.

THE FINAL CONSONANT. To synthesize the final consonant, I selected a representative example from the recorded non-words [skap] and [spak]. I selected a token in which the final stop had an audible release that was also visible on the spectrogram and in the waveform. I spliced the final stop from the two representative tokens, and stored it digitally as the *k*-file and the *p*-file respectively. A 20-step continuum ranging from [k] to [p] was then created by proportionally mixing together the *k*-file and *p*-file. The first token on the continuum consisted of only the *k*-file and the last token of the only *p*-file. All of the intermediate tokens consisted of a blending of the *k*-file and *p*-file in different proportions.

PUTTING THE PIECES TOGETHER. The [k] to [p] continuum was then spliced onto each of the nucleus continua. A silence (consonantal closure) of 65 ms was included between the end of the nucleus and final consonant.

Finally, a representative example of the onset sequence [sk] was selected from the recordings of [skap], and a representative example of [sp] from the recordings of [spak]. The [sk]-onset was spliced onto each of the tokens in the [skak] to [skap] continuum, and the [sp]-onset onto each of the tokens in the [spak] to [spap] continuum. The final result is a 20-step continuum from [skak] to [skap], and 20-step continuum from [spak] to [spap]. The other continua were constructed in a similar way. This resulted in tokens of high quality that sounded like natural speech.

5.4. SUBJECTS

All subjects were undergraduate students from the University of Massachusetts, who took part in the experiment for extra course-credit. Subjects were native speakers of American English, and reported no hearing disabilities. A total of 37 subjects took part. In each condition only those subjects who could identify the endpoint stimuli correct at least 75% of the time were included in

the analysis. This resulted in 15 subjects in the K~P-condition, 26 in the K~T-condition, and 26 in the P~T-condition. Subjects were tested either individually, or in groups of up to four.

5.5. PROCEDURE

A separate experiment file was created for each of the three conditions. The K~P-condition is discussed as an example. The K~P-file contained three repetitions of each of the four endpoint tokens ([skak], [skap], [spap], [spak]). Of the 18 intermediate tokens from each continuum, six were selected for inclusion in the experiment. These six tokens were selected so that they surrounded the most ambiguous region on the continuum.¹⁵ Six repetitions of each of the six ambiguous tokens for each of the two continua were included in the K~P-file. Filler items from a [trap]~[trak]-continuum were also included. Files for the other conditions were created in the same way, with fillers taken from a [kraʌp]~[kraʌt]-continuum [præk]~[pret]-continuum.

The three conditions were tested separately. Stimuli were presented at a comfortable level over headphones to subjects seated in a sound attenuated room. Subjects pressed a button corresponding to the sound on which the token ended. The tokens in each file were presented in random order. Before an experiment file was presented, it was preceded by 18 practice trials (three repetitions of each of the six endpoint tokens) without feedback.

5.6. RESULTS

In each condition, there was clear evidence of a perceptual bias in the expected direction. In the K~P-condition, the average percent [p]-responses on the [spap]~[spak]-continuum was 39%, and on the [skap]~[skak]-continuum it was 48%, corresponding to a bias against [p] on the [spap]~[spak]-continuum and towards [p] on the [skap]~[skak]-continuum (one-tailed $t(14) = 1.712$, $p = .055$). The mean identification functions for these two continua are shown in Figure 1a.

In the T~K-condition, the percent [k]-responses were calculated on each continuum. Here we expect a bias against [k] on [skæk]~[skæt]-continuum because [skæk] would violate *skVk. However, on the [stek]~[stet]-continuum, we expect a bias toward [k] only if [stet] does indeed violate *stVt. Either way, more [k]-responses are expected on the [stek]~[stet]-continuum than on the [skæk]~[skæt]-continuum. What is not clear yet is whether this is due to a bias on only one of the continua or on both of the continua. A bias in the expected direction was indeed found. The average percent [k]-responses on the [skæk]~[skæt] continuum was 47%, and on the [stek]~[stet]-continuum 55% (one-tailed $t(25) = 2.822$, $p < .005$). These results are represented in Figure 1b.

¹⁵ These six tokens were identified as follows: In a pre-test, a 20-step [hap]~[hak]-continuum was synthesized. The tokens on this continuum were identical to the tokens on the [skap]~[skak]-continuum and the [spap]~[spak]-continuum in the last 60% of the nucleus and the final stop. They differed only in that they started with the place neutral consonant [h-]. This continuum was presented to 20 subjects in a pre-test. Subjects had to identify the final consonants as either [p] or [k]. Those six tokens that surrounded the 50%-mark on this continuum were assumed to be the six most ambiguous tokens.

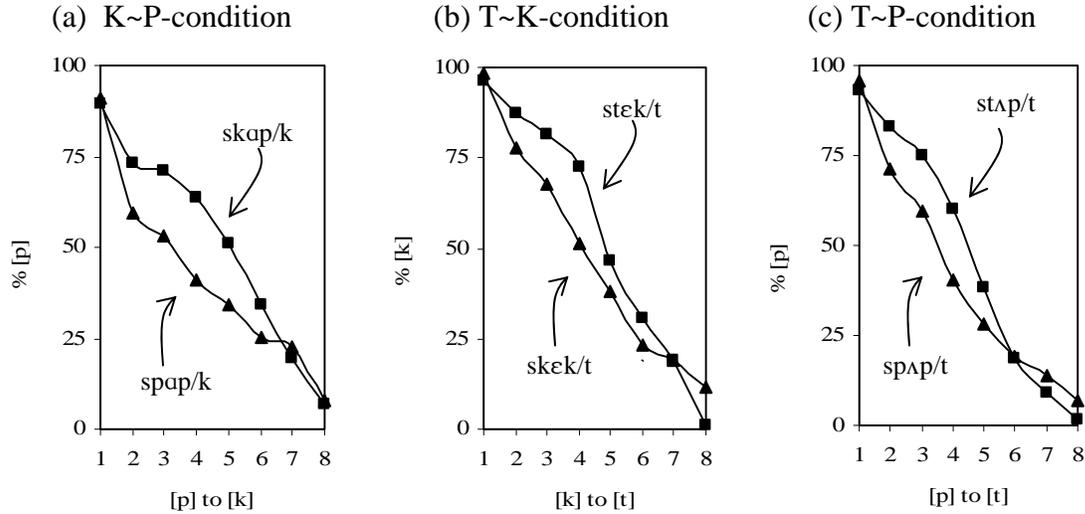


FIGURE 1. Identification functions for the three conditions in the phoneme identification experiment

In the T~P-condition, I measured the percent [p]-responses. Here we expect more [p]-responses on the [stap]~[stat]-continuum than on the [spap]~[spat]-continuum. However, as in T~K-condition, it is not clear yet whether this is due to just one constraint or two. The expected bias was found – the average percent [p]-responses on the [spap]~[spat]-continuum was 40%, and on the [stap]~[stat]-continuum 47% (one-tailed $t(25) = 1.897$, $p < .04$). These results are represented in Figure 1c.

If *stVt is not part of the grammar of English, then the bias-causing constraints in the K~P-condition (*spVp, *skVk) form a proper superset of the bias-causing constraints in the T~K-condition (*skVk) and also in T~P-condition (*spVp). We would then expect a smaller bias-effect in the T~K-condition and the T~P-condition than in the K~P-condition. However, if *stVt is indeed part of the grammar of English, then there are two bias-causing constraints in the K~P-condition (*spVp, *skVk), in the T~K-condition (*skVk, *stVt), and in the T~P-condition (*spVp, *stVt), and we would not necessarily expect a larger bias-effect in the K~P-condition. The size of the bias effect in the three conditions is indeed very similar – see Table 2. To test whether the effect sizes differ between the conditions, I computed a difference score for each subject. For instance, in the K~P-condition, I subtracted the percent [p] responses of each subject on the [spap]~[spak]-continuum from his/her percent [p]-responses on the [skap]~[skak]-continuum. Two tailed paired sample t -tests were performed on each of the three pairs of difference scores. (I included only subjects with a difference score in each of the two conditions being compared in these calculations.) No significant difference was found; that is, there is no evidence that the bias size is different between the three conditions. The values obtained are: (i) P~K- and P~T-conditions: $t(13) = 0.403$, $p = 0.69$; (ii) P~K- and K~T-conditions: $t(10) = 0.748$, $p = 0.47$; (iii) P~T- and K~T-conditions: $t(21) = 0.829$, $p = .42$.

Condition	Continuum	Response pattern	Bias size
K~P	[spap]~[spak]	39%	9%
	[skap]~[skak]	48%	
T~K	[skɛk]~[skɛt]	47%	8%
	[stɛk]~[stɛt]	55%	
T~P	[spʌp]~[spʌt]	40%	7%
	[stʌp]~[stʌt]	47%	

TABLE 2. The size of the perceptual bias in terms of the total response pattern

There is no evidence that the size of the perceptual bias in the T~K- and T~P-condition is different from the size in the K~P-condition. We know that the bias in the K~P-condition is the result of a biasing constraint on each of the two continua. This indicates that there should also be a biasing constraint on each of the continua in the T~K- and T~P-conditions, and therefore that [stɛt] and [stʌt] do violate *stVt. When given a choice between an [stVt]-form and either an [stVp]- or [stVk]-form, English listeners prefer the [stVp]- and [stVk]-forms. Although [stVt]-forms are possible words of English, they are therefore not perfect.

5.7. FREQUENCY STATISTICS

In the selection of stimuli, the endpoint tokens were equalized in terms of their TP's. However, this was not possible for LND. To confirm that the LND difference between the endpoints was not the cause of the perceptual biases, I performed a regression analysis on the response data, using the LND of the tokens as the independent variable. To understand how this analysis was done, consider the K~P-condition as an example. The LND for [spak] is 35.69 and [spap] it is 21.41. (See the appendix for the frequency statistics associated with all of the stimuli.) The LND of the [k]-endpoint is 14.28 higher than that of the [p]-endpoint. This means that there is negative bias of 14.28 against the [p]-percept on the [spap]~[spak]-continuum. On the [skap]~[skak]-continuum, the neighborhood bias is in favor of [p]. The neighborhood density of [skap] is 31.57, and of [skak] 18.41. The difference between these two translates into a 13.16 positive bias towards [p]. I did similar calculations for the other conditions. A regression analysis can now be done using these values as the independent variable. The percent [p]-responses in the K~P-condition, the percent [k]-responses in the T~K-condition, and the percent [p]-responses in the T~P-condition for each subject are used as dependent variable. The r^2 -value attained for such an analysis is 0.065. The LND differences account for only a small fraction of the variation observed in the data.

The two frequency statistics that have been shown in the literature to influence speech processing, LND and TP, can therefore not fully account for the response pattern observed in this experiment. However, frequency is a very general concept and many different kinds of frequencies can be calculated over the lexicon. It is hence possible that some frequency statistics other than LND or TP are better correlated with the response patterns. Although it is impossible to show that this is not the case for every conceivable way of calculating frequency over the

lexicon, I consider two likely alternative statistics here. The main point of this experiment is that there exists a perceptual bias against [t], due to its violation of *stVt. Berkley (2000) has shown that words with coronals in onset and coda position are underrepresented in the English lexicon relative to words with a coronal and labial/dorsal in onset and coda position. It is therefore possible that there are just more sequences of the form $\sigma[.t.V.p...]\sigma$ or $\sigma[.t.V.k...]\sigma$ than $\sigma[.t.V.t...]\sigma$ in English, and that the perceptual bias might come from this. Similarly, it is possible that there are more sequences of the form $\sigma[stVp]\sigma$ or $\sigma[stVk]\sigma$ than $\sigma[stVt]\sigma$. I calculated the frequencies of these syllable types in the CELEX syllable sub-corpus. The results are given in Table 3. Syllables with two [t]'s are more frequent than syllables with [t] and [p]/[k]. If frequency counts such as these were the cause of the results of this experiment, a bias towards rather away from [t] should have been observed.

Place	Syllable type	Frequency	Syllable type	Frequency
Coronal	$\sigma[.t.V.t...]\sigma$	5517	$\sigma[stVt]\sigma$	910
Dorsal	$\sigma[.k.V.k...]\sigma$	4414	$\sigma[skVk]\sigma$	587
Labial	$\sigma[.p.V.p...]\sigma$	1767	$\sigma[spVp]\sigma$	304

TABLE 3. Syllable frequencies per million words in CELEX

6. DEGREES OF UNGRAMMATICALITY I: WORD-LIKENESS EXPERIMENT

In this experiment, English listeners were presented with non-word pairs of three kinds: [stVt]~[skVk], [stVt]~[spVp], or [skVk]~[spVp]. The task of subjects was to select from each pair the member that they considered to sound most like it could be a word of English.

6.1. TOKEN SELECTION

As with phoneme identification, it is known that frequency statistics influence word-likeness rating – non-words with high values in terms of frequency statistics are typically rated as more word-like (Bailey and Hahn 1998, 2001; Coleman and Pierrehumbert 1997; Frisch et al. 2001). In order to control for this, tokens were selected as far as possible so that grammar and frequency statistics conflict – i.e. the member of a pair that is less well-formed according to the grammatical analysis in section 4 above had higher values in terms of frequency statistics. If we find that the subjects respond according to the predictions of the grammatical analysis, we can therefore be rather certain that the results come from the grammar and not from the frequency statistics. I calculated the lexical neighborhood density (LND) and the cumulative bi-phone probability (CBP) for each token (CBP is the product of all the individual bi-phone probabilities for a token). One-tailed, paired sample *t*-tests confirmed that the LND and CBP of the less well-formed members were significantly higher in all three conditions. A list of all the tokens used and their associated frequency statistics is included in the appendix.

6.2. SUBJECTS

The subjects were 20 undergraduate students from the University of Massachusetts. All subjects were native speakers of American English, and reported no speech or hearing disabilities. Subjects took part in the experiment for course credit in an introductory Linguistics class.

6.3. PROCEDURE

I recorded a phonetically trained native female speaker of American English reading each of the tokens and some filler items in the carrier sentence ‘John said _____ again to me.’ The tokens and fillers were then sliced from the carrier sentences, and were digitally equalized for intensity.

There were 15 token-pairs in each of the three conditions, resulting in 45 test-pairs. I added 45 filler pairs to this. The fillers were selected such that approximately an equal number of all tokens were possible words and impossible words. Fillers that represented impossible words violated a constraint on the consonants that co-occur in the onset and coda of a single syllable (Fudge 1969). They were therefore ill-formed for reasons similar to the ill-formedness of the [skVk]- and [spVp]-tokens. This resulted in a total of 90 token-pairs. Two lists were created from these 90 token-pairs. Each list contained all 90 token-pairs. In List 1, eight out of the fifteen pairs of the T~K-condition had the [stVt]-token first and the [skVk]-token second. In the other seven token-pairs for this condition, the [skVk]-token was used first. The same was true for the T~P-pairs and K~P-pairs. In List 2, the order between the members in a token pair was reversed. Both lists were presented auditorily to subjects over headphones. About 5 minutes elapsed between the presentation of the lists. On each presentation of a list, it was differently randomized. Before the list was presented the first time, 10 filler token-pairs were presented as practice trials. Subjects were seated in a sound-attenuated room, and were tested individually or in groups of up to four.

Subjects were instructed that they would hear a pair of non-words, and that their task would be to select the member of each pair that sounded most like it could be a word of English. Subjects indicated their response by pushing one of two buttons. The left-hand button was pushed if the first token in a pair was preferred, and the right-hand button if the second token was preferred. The next token pair was presented 500 ms after all subjects had responded, or after 5 seconds had elapsed.

6.4. RESULTS

In each test-pair, there was a member that is more ill-formed according to the grammatical analysis presented in section 4. In the K~P-condition, this was [spVp]; in the T~K-condition, it was [skVk]; and in the T~P-condition, it was [spVp]. The results of the experiment were scored as follows: If a subject was presented with the token-pair [Token 1]~[Token 2], and he/she chose [Token 1], then [Token 1] was assigned a score of 1 and [Token 2] a score of 0. The results were submitted to a 2 × 3 ANOVA (well-formedness × condition). A main effect of well-formedness

was found both by subjects ($F(1,19) = 23.28, p < 0.001$) and by items ($F(1,14) = 188.43, p < 0.001$). There was also a significant interaction between well-formedness and condition by subjects ($F(2,18) = 10.37, p = 0.001$) and items ($F(2,13) = 23.91, p < 0.001$). The results are represented in Figure 2.¹⁶

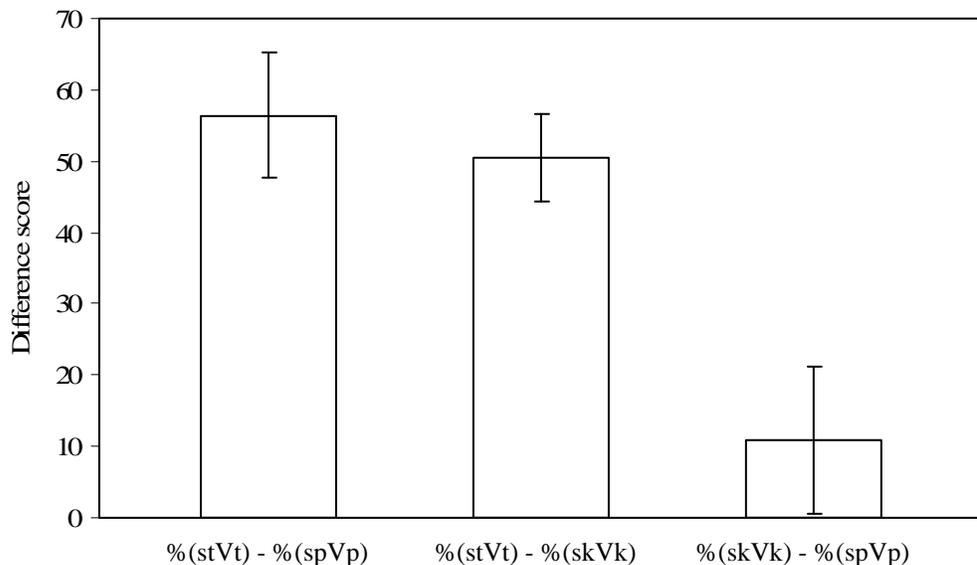


FIGURE 2. Response patterns in word-likeness experiment

Since the main interest of this experiment was to test for a difference between [skVk] and [spVp], this contrast was investigated further with paired sample *t*-test. There was a general advantage for the more well-formed [skVk]-tokens over the less well-formed [spVp]-tokens, that was significant by items ($t(14) = 1.94, p < 0.037$), but not subjects ($t(19) = 1.00, p = 0.16$). Although the subjects comparison did not reach significance, the significant result for the items analysis confirms that the well-formedness difference between [skVk] and [spVp] forms is reflected in the word-likeness judgments.¹⁷

The results of this experiment confirm that grammar does influence word-likeness judgments. When given a choice between two forms that differ in well-formedness, language users are biased to select the more well-formed token, even if both forms are impossible words. Assuming that this result originates in grammar, we can take this as evidence for the ranking $*spVp \gg *skVk$. The kind of data traditionally used in phonology would not have been able to give evidence for this ranking. Traditional data would only have stated that neither [skVk] nor

¹⁶ The bars represent difference scores calculated for each condition by subtracting the percent times that the less well-formed member of a pair was selected from the percent times that the more well-formed member was selected. A positive difference score therefore indicates that the more well-formed member was selected more often. Error bars mark the 95% confidence interval.

¹⁷ The contrasts in the other two conditions, though not the focus of this experiment, were also investigated. There was significant advantage for [stVt] over [skVk] (subject $t(19) = 4.5, p < .001$; item $t(14) = 5.5, p < .001$), and over [spVp] (subject $t(19) = 5.7, p < .001$; item $t(14) = 5.4, p < .001$). This is not surprising, as the comparison here is between a possible and an impossible word.

[spVp] is a possible word of English. This would have been evidence that *skVk and *spVp outrank faithfulness, but not for any ranking between these two constraints.

6.5. FREQUENCY STATISTICS

It is significant to note that the response pattern in this experiment is in direct conflict with what would be expected based on the TP's and LND's of the tokens. In each condition, the tokens that were selected less frequently by the subjects had higher values in terms of frequency statistics. This experiment therefore serves as strong evidence that in addition to frequency statistics, grammar also contributes to word-likeness ratings. To confirm that the response pattern cannot be explained by the frequency statistics, regression analyses were performed on the results using the LND's and the CBP's of the tokens as the independent variables. The r^2 -values for both of these analyses were smaller than 0.03. This confirms that the frequency statistics cannot account for the observed variation in the response data.

Having shown that the frequency statistics that are known from the literature to influence non-word processing do not contribute significantly to the results of this experiment, it is unlikely that any other kind of frequency statistic would. However, as noted with regard to the previous experiment, frequency is a very broad concept and there may be another kind of frequency statistic that correlates better with the response pattern. Specifically, if the type frequency of $\sigma[.k...k.]_{\sigma}$ is higher than that of $\sigma[.p...p.]_{\sigma}$, then the result might come from this. As with the previous experiment, I calculated the frequency of syllables of the form $\sigma[.t...t.]_{\sigma}$, $\sigma[.k...k.]_{\sigma}$ and $\sigma[.p...p.]_{\sigma}$ in CELEX's syllable sub-corpus. $\sigma[.t...t.]_{\sigma}$ has a frequency of 5517, $\sigma[.k...k.]_{\sigma}$ of 745, and $\sigma[.p...p.]_{\sigma}$ of 228. As can be seen, it is indeed the case that $\sigma[.k...k.]_{\sigma}$ is more frequent than $\sigma[.p...p.]_{\sigma}$, and it is hence at least possible that this difference could explain the results of the experiments. To test whether this is the case, I took the natural logarithm of the frequency of each of $\sigma[.t...t.]_{\sigma}$, $\sigma[.k...k.]_{\sigma}$ and $\sigma[.p...p.]_{\sigma}$, and performed a regression on the response patterns of the experiment using these transformed frequency counts as independent variables. This returned an r^2 -value of 0.14, which, though higher than the result obtained for the LND and CBP regressions, is still rather low. There is still a large amount of the variation in the data that cannot be explained merely on the grounds of this frequency.

Since the kind of statistic discussed in the previous paragraph has never been shown to influence word-likeness judgments, since the correlation between this statistic and the results are rather low, and since the frequency statistics that are known to influence word-likeness judgments (LND and CBP) correlate even less well with the results, I conclude that it is very unlikely that the response patterns observed in this experiment can be explained solely on the grounds of frequency statistics.

7. DEGREES OF UNGRAMMATICALITY II: LEXICAL DECISION EXPERIMENT

In this experiment, I presented subjects auditorily with a list of words and non-words, with the task to discriminate between words and non-words. The hypothesis is that listeners use, among other things, the information provided by grammar when they make lexical decisions. The less well-formed a non-word token is, the less seriously a listener will consider it as a possible word,

and the more quickly the token will be rejected (Berent, Everett and Shimron 2001; Berent, Shimron and Vaknin 2001; Berent et al. 2002; Kager and Shatzman 2007).

7.1. TOKEN SELECTION

This experiment had the same three conditions as the word-likeness experiment – i.e. T~K, T~P, and K~P. Consider the T~K-condition as an example. I selected 5 non-words of the form [stVt] and 5 non-words of the form [skVk]. Lexical decision is also influenced by frequency statistics (Balota and Chumbley, 1984; Luce and Pisoni, 1998; Vitevitch and Luce, 1999). Non-words with high LND's and high TP's are typically identified as non-words more slowly. For this reason, the tokens were selected such that the mean LND and the mean CBP of the [stVt]-tokens and the [skVk]-tokens did not differ significantly. Frequency statistics therefore did not favor any of the two token types. The token selection in the other two conditions was done in a similar way. The actual tokens used, as well the frequency statistics associated with them, are given in the Appendix.

7.2. SUBJECTS

The same 20 subjects who took part in the word-likeness experiment also took part in the lexical decision experiment. Half of the subjects participated in the word-likeness judgment experiment first, and the other half in the lexical decision experiment.

7.3. PROCEDURE

I recorded the same speaker as in the word-likeness experiment reading each of the tokens as well as some filler items in the carrier sentence 'John said _____ again to me.' Fillers were selected in the same manner as for the word-likeness experiment. The tokens and fillers were sliced from the carrier sentence and digitally equalized for intensity. The stimulus list that was presented to subjects was structured as follows: There were a total of 27 test-tokens.¹⁸ Each of these 27 test-tokens was included once in the list. I added 76 monosyllabic fillers to the list so that the final list contained 103 tokens. Fillers were selected so that the final list contained roughly the same number of actual words and non-words. The non-word fillers were selected so that half of all non-words tokens (fillers and test tokens) were possible words of English and half were not possible words. [stVt] non-words, as well as forms like [blɪk] counted as possible non-words. Like the [spVp] and [skVk] test-tokens, the fillers that counted as impossible words were impossible because they contained disallowed combinations of consonants in onset and coda – e.g. forms like [klæɪ], [smɪŋ], etc. These tokens were presented to subjects over headphones. On each presentation the list was differently randomized. The list was presented twice to subjects, with a break of about five minutes between presentations. Subjects responded by pressing one of

¹⁸ This number is smaller than the expected number of 30 (3 conditions × 2 token types per condition × 5 tokens per token type), because the same token is sometimes used in two different conditions.

two buttons on a response box. One button was marked as ‘Yes’, and was used to indicate that the token was a word of English. The other button was marked as ‘No’, and was used to indicate that the token was not a word of English. The order between the buttons was varied so that half of the subjects responded ‘Yes’ with their right hands, and half responded ‘No’ with their right hands. Subjects were instructed to respond as quickly as possible, but to listen to the whole token before responding. The next token was presented after all subjects have responded or after 2 seconds have elapsed. Both responses and response times were recorded. Before the list was presented the first time, 10 filler tokens were presented as practice trials. Subjects were seated in a sound attenuated room, and were tested individually or in groups of up to 4.

7.4. RESULTS

Response times were recorded starting at the onset of a stimulus. Before analysis, the duration of every stimulus was subtracted from the recorded response time. The resulting measure represented how long after (of before) the end of the stimulus a subject recorded a response. In what follows, I refer to this measure as ‘response time’. Tokens with negative response times, indicating that the subject responded before the end of the stimulus, were excluded from further analysis. The remaining response times for each subject were *Z*-score transformed, and responses that were more than 2 standard deviations away from the mean for a subject were excluded from the analysis. Only correct non-word responses were included in the analysis. Exclusion of negative response time tokens, outliers and incorrect responses resulted in the exclusion of just over 8% of the total responses. The results are represented graphically in Figure 3.¹⁹

In each condition, one of the token types were more marked according the grammatical analysis presented in section 4. The response time data were subjected to a 2×3 ANOVA with markedness (marked, unmarked) and condition (K~P, T~K, T~P) as independent variables. A main effect of markedness was found by subjects ($F(1, 19) = 21.68, p < 0.001$), but not by items ($F(1, 4) = 2.584, p = 0.18$). There was no interaction between markedness and condition by subjects ($F(2, 18) = 0.58, p = 0.57$) or by items ($F(2, 3) = 0.08, p = 0.93$). Since the main interest of this experiment was to test for a reaction time difference between [skV_k]- and [spV_p]-forms, this contrast was investigated further. On average, [spV_p]-tokens were identified as non-words faster than [skV_k]-tokens. This difference was significant both by items (one tail, two sample $t(8) = 2.15, p = .03$) and subjects (one tail, paired samples $t(19) = 3.79, p < .001$).²⁰

¹⁹ A difference score was calculated for each subject in each condition, by subtracting the response time to the less well-formed tokens from the response time to the more well-formed tokens. A positive difference score therefore indicates that the more well-formed tokens were rejected as non-words more slowly. Error bars mark the 95% confidence interval.

²⁰ The contrasts in the other two conditions, though not the focus of this experiment, were also investigated. [stV_t] forms were rejected more slowly than [skV_k] and [spV_p]. This difference was significant by subjects (T~K: $t(19) = 3.4, p < .002$; T~P: $t(19) = 4.2, p < .001$) and tended towards significance by items (T~K: $t(8) = 1.6, p = .07$; T~P: $t(8) = 1.6, p = .07$).

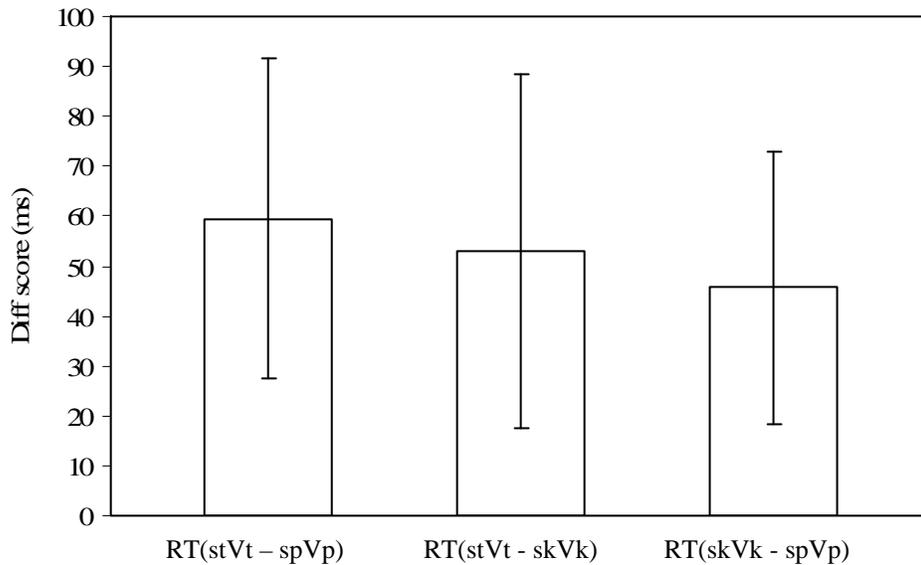


FIGURE 3. Response time differences between token types in lexical decision experiment

These results confirm the general hypothesis about how grammar influences lexical decision. The less well-formed [spVp]-tokens were rejected more quickly than the more well-formed [skVk]-tokens. This serves as evidence of the ranking $*spVp \gg *skVk$, and it therefore confirms that English grammar makes well-formedness distinctions even between ungrammatical forms. Again we see how data from psycholinguistic experiments can help us to test hypotheses about the grammar of English that we would not have been able to test had we relied only on the kinds of data used more traditionally in phonology.

7.5. FREQUENCY STATISTICS

The tokens for this experiment were selected such they were about equal in terms of LND and CBP. We therefore do not expect these statistics to make a significant contribution to the response patterns. Using the LND's and CBP's of the tokens as independent variables, I performed regression analyses on the response time data. The corresponding r^2 -values are 0.02 and 0.01. These frequency statistics therefore account for only a small fraction of the variation observed in the data. As with the word-likeness experiment, I also performed a regression using the log frequencies of syllables of the form $\sigma[.t\dots t.]_{\sigma}$, $\sigma[.k\dots k.]_{\sigma}$ and $\sigma[.p\dots p.]_{\sigma}$ as the independent variables. This regression returned an r^2 -value of 0.03. Although there are undoubtedly more frequency statistics that can be calculated, given the poor correlation between the statistics discussed here and the reaction times, it is unlikely that any other frequency statistic will correlate very well with the reaction times. I therefore conclude that the results of this experiment cannot be ascribed solely to the influence of frequency statistics.

8. CONCLUDING REMARKS.

8.1. FREQUENCY AGAIN

Under a strictly usage-based approach to language 'grammar is the cognitive organization of one's experience with language' (Bybee 2006:711). If this view is taken to its extreme, then

'phonological grammar is a simple projection of lexical statistics' (Hay et al. 2003:59). On the other extreme are views that reject the role of language usage and frequency information in grammar, relegating these aspects to performance rather than competence, and hence placing it outside of the purview of grammar. Although there are probably no, or very few, linguists who subscribe to either of these extremes, much of recent linguistic research has focused on finding evidence that formal grammar to the exclusion of usage frequencies influence language processing or *vice versa*. (The current paper also falls into one of these categories.) Consequently, there is a growing literature that gives evidence for both of these influences on linguistic processing. However, what is still largely lacking is research that assumes that both of these influence processing and that asks how they interact. What happens if usage frequency statistics and formal grammar conflict? Are there certain kinds of tasks that are more susceptible to the influence of frequency statistics, and others that are more susceptible to the influence of formal grammar? Where does formal grammar fit into the language processing models developed by psycholinguists? The time has come to end the debate between the two extremes, to accept that both frequency statistics and formal grammar play a role, and to investigate the relation between these two.

Although research that asks these kinds of questions is not plentiful, it would be untrue to say that no research on this question has been done. Frisch and Zawaydeh (2001) is a good example of research that addresses exactly this question. They take the stance that grammar is 'emergent' from the lexicon, by which they mean that grammar is learned 'based on lexical patterns' (2001:104-105). However, at the same time, they found that the response patterns of the subjects in their experiment could not be explained solely on the basis of the frequency-based lexical patterns, and they hence conclude that 'knowledge of phonotactic constraints is independent from simple lexical statistics' (2001:104). Grammar is learned from or emerges from exposure to language usage. This is in basic agreement with Bybee's definition of grammar as 'the cognitive organization of one's experience with language' (2006:711). However, once this organization has happened, it gains a level of independence from this experience so that it can generalize into areas where the language learner has not had any experience.

I subscribe to a similar view of the relation between formal grammar and usage frequencies. The grammar of a specific language is acquired by users of that language only through exposure to use of that language. This experience has a rich frequency structure, and language users can and do use this frequency information to acquire the grammar of the language. This is in agreement with much of the formal learnability research in phonology (Boersma 1998; Boersma and Hayes 2001; Hayes and Wilson to appear; Pater 2005). I also showed in section 4.3 how English learners can acquire the ranking $*p_v p \gg *k_v k \gg *t_v t$ only based on usage frequency statistics. However, once this part of the grammar has been learned, it becomes partially independent from the usage statistics and is generalized to domains with which English users have no experience. Specifically, they infer from this learned ranking the ranking $*spVp \gg *skVk$. Crucially, the absence of words of the form [spVp] and [skVk] from English means that this ranking cannot be learned from usage frequencies. Yet, we saw clear evidence for this ranking in both the word-likeness and the lexical decision experiments discussed in sections 6 and 7 above.

The way in which formal grammar and usage frequency statistics interact in language processing still needs to be researched in much more detail. New research should no longer try to

show that the one or the other influences processing, but should accept that both do, and investigate the way in which they interact.

8.2. THE PROPERTIES OF AN ADEQUATE GRAMMATICAL MODEL

The results of the three experiments reported here, together with other similar results from the literature, teach us something about the nature of grammar in general. They show that the grammar of an individual language retains aspects of Universal Grammar that are not active in the grammar of the individual language (*stVt is part of English grammar although it is not necessary). Also, the grammars of individual languages have more structure than what is necessary to distinguish between grammatical and ungrammatical forms (English has the ranking *spVp \gg *skVk although this ranking is not necessary).

This tells us something about the requirements that any model of grammar must meet in order to account adequately for the grammatical competence of language users. A strictly generative grammar is not adequate. Such a grammar can distinguish between grammatical and ungrammatical – a grammatical form is a form for which a well-formed derivation exists and all other forms are ungrammatical. However, such a grammar cannot compare different grammatical forms with each other, nor can it compare different ungrammatical forms with each other. A strictly generative grammar lacks the necessary comparative abilities. An adequate model of grammatical competence will be able to make the distinction between grammatical and ungrammatical. But in addition to this, it will also be able to compare grammatical forms with each other in terms of their well-formedness, and ungrammatical forms with each other in terms of their ill-formedness.

OT is a prime example of a grammatical model with exactly these properties. An OT grammar consists of a generative component (GEN) that takes an input /in/, and creates a set of output candidates. These candidates are then subjected to the evaluative component (EVAL) that compares them and selects the best candidate as the output. We can then represent the grammar G as a function that takes the input /in/ as argument and returns the output [out] as follows: $G(/in/) = \text{EVAL}(\text{GEN}(/in/)) = [\text{out}]$ (McCarthy 2002:10). The grammar is therefore generative in nature – it still maps an input onto its unique grammatical output. However, an OT grammar also has a comparative component, EVAL. An OT grammar can therefore make the distinction between grammatical and ungrammatical. But it can also compare grammatical forms with each other, and ungrammatical forms with each other. See Coetzee (2004, 2006) and Berent and Shimron (1997) for the formal development of both of these properties of an OT grammar.

The fact that OT has these properties does not mean that it is *the* correct model of grammar. Other models of grammar can also be developed that will have these two properties. And in fact, there are other models with these properties that can probably account even better for results of the kinds of experiments discussed here. A prime example is Harmonic Grammar (Smolensky and Legendre 2006; Pater et al. 2007). Harmonic Grammar preserves the comparative character of OT. However, unlike classic OT where constraint ranking is discrete so that only relative well-formedness can be gleaned from an OT grammar, constraints in Harmonic Grammar can take on any weight so that candidates can receive an absolute numeric goodness value. The details of Harmonic Grammar still need to be worked out. However, it has the promise to bring important new advances to our understanding phonological grammar and how it interacts with phonological processing.

APPENDIX: TOKENS USED IN EXPERIMENTS

TOKENS USED IN THE PHONEME IDENTIFICATION EXPERIMENT

Condition	Endpoint	LND	Transitional Probability	
			Probability of V given final C	Probability of final C given V
K~P	[spap]	21.41	0.79	0.76
	[spak]	35.69	0.78	0.73
	[skap]	31.57	0.79	0.76
	[skak]	18.41	0.78	0.73
T~K	[skæk]	12.65	0.82	0.82
	[skæt]	34.37	0.79	0.81
	[stæk]	47.58	0.82	0.82
	[stæt]	38.11	0.79	0.81
T~P	[spʌp]	11.38	0.76	0.79
	[spat]	31.13	0.81	0.76
	[stʌp]	46.08	0.76	0.79
	[stat]	42.19	0.81	0.76

TOKEN PAIRS USED IN WORD-LIKENESS JUDGMENT EXPERIMENT

T~K-CONDITION

[stVt]	LND	CBP		[skVk]	LND	CBP
stɔ:t	23.06	0.217		skæk	22.12	0.227
stət	17.06	0.144		skɑ:k	18.41	0.219
stɔɪt	10.65	0.066		skæk	12.65	0.193
stɔɪt	10.65	0.066		skɑ:k	18.41	0.219
stɔɪt	10.65	0.066		skʌk	19.38	0.183
stɔɪt	10.65	0.066		skæk	22.12	0.227
stət	17.06	0.144		skɛk	26.06	0.186
stʌt	42.19	0.219		skɪk	37.31	0.299
stɔ:t	23.06	0.217		skɪk	37.31	0.299
stət	17.06	0.144		skɪk	37.31	0.299
stət	17.06	0.144		skæk	22.12	0.227
stɔɪt	10.65	0.066		skark	13.89	0.157
stɔɪt	10.65	0.066		skɛk	26.06	0.186
stət	17.06	0.144		skæk	12.65	0.193
stɔɪt	10.65	0.066		skɪk	37.31	0.299
<i>Average</i>	16.54	0.122		<i>Average</i>	24.21	0.228

Paired sample, one-tailed *t*-test on LND: $t(14) = 3.32, p = .005$, and CBP: $t(14) = 7.27, p < .001$.

T~P-CONDITION

[stVt]	LND	CBP		[spVp]	LND	CBP
stɔ:t	23.06	0.217		spɪp	26.69	0.212
stɔ:t	10.65	0.066		spʌp	11.38	0.182
stɔ:t	10.65	0.066		spi:p	27.45	0.213
stɔ:t	10.65	0.066		spɪp	26.69	0.212
stɔ:t	23.06	0.217		spi:p	27.45	0.213
stɔ:t	10.65	0.066		spɑ:p	21.41	0.244
stɔ:t	10.65	0.066		spæp	17.98	0.201
stut	17.06	0.144		spæp	17.98	0.201
stut	17.06	0.144		spɪp	26.69	0.212
stɔ:t	10.65	0.066		speɪp	17.13	0.163
stɔ:t	10.65	0.066		sep	20.32	0.208
stut	17.06	0.144		spɑ:p	21.41	0.244
stut	17.06	0.144		spi:p	27.45	0.213
stut	17.06	0.144		sep	20.32	0.208
stut	17.06	0.144		spʌp	11.38	0.182
<i>Average</i>	14.87	0.117		<i>Average</i>	21.45	0.207

Paired sample, one-tailed t -test on LND: $t(14) = 4.30, p < .001$, and CBP: $t(14) = 6.28, p < .001$.

K~P-CONDITION

[skVk]	LND	CBP		[spVp]	LND	CBP
skauk	9.95	0.001		spæp	17.98	0.201
sku:k	12.64	0.193		spæp	17.98	0.201
skauk	9.95	0.001		spɑ:p	21.41	0.244
sku:k	12.64	0.193		spɑ:p	21.41	0.244
skuk	3.86	0.097		spɑ:p	21.41	0.244
skauk	9.95	0.001		sep	20.32	0.208
sku:k	12.64	0.193		sep	20.32	0.208
skuk	3.86	0.097		sep	20.32	0.208
ski:k	28.63	0.134		spɪp	26.69	0.212
skɔ:k	10.04	0.171		spɪp	26.69	0.212
skuk	3.86	0.097		spɪp	26.69	0.212
skɔ:k	10.04	0.171		spi:p	27.45	0.213
skuk	3.86	0.097		spi:p	27.45	0.213
skauk	9.95	0.001		spʌp	11.38	0.182
sku:k	12.64	0.193		spʌp	11.38	0.182
<i>Average</i>	10.30	0.109		<i>Average</i>	21.26	0.212

Paired sample, one-tailed t -test on LND: $t(14) = 5.26, p < .001$, and CBP: $t(14) = 5.01, p < .001$.

TOKENS USED IN LEXICAL DECISION EXPERIMENT

K~P-CONDITION

[sKvK]	LND	CBP		[sPvP]	LND	CBP
ska:k	13.89	0.157		spa:p	30.94	0.398
ska:uk	9.95	0.001		spa:up	5.20	0.006
ska:k	18.41	0.219		spa:p	21.41	0.244
ski:k	28.63	0.134		spo:p	2.38	0.039
sku:k	3.86	0.097		spa:p	11.38	0.182
<i>Average</i>	14.95	0.121		<i>Average</i>	14.26	0.174

Two tailed, two sample *t*-tests on LND: $t(8) = 0.10, p > .92$, and CBP: $t(8) = 0.06, p > .95$.

T~K-CONDITION

[sTvT]	LND	CBP		[sKvK]	LND	CBP
sto:t	23.06	0.217		sker:k	26.06	0.186
start	40.01	0.303		ski:k	28.63	0.133
stet	38.11	0.285		skik	37.31	0.299
stu:t	39.66	0.335		skæk	13.89	0.227
stot	17.06	0.144		sku:k	22.12	0.193
<i>Average</i>	31.58	0.257		<i>Average</i>	25.60	0.208

Two tailed, two sample *t*-tests on LND: $t(8) = 0.78, p > .45$, and CBP: $t(8) = 1.81, p > .11$.

T~P-CONDITION

[sTvT]	LND	CBP		[sPvP]	LND	CBP
sto:t	23.06	0.217		spa:p	21.41	0.244
start	40.01	0.303		spa:p	30.94	0.398
stet	38.11	0.285		spi:p	27.45	0.213
stu:t	39.66	0.335		spi:p	26.69	0.212
stot	17.06	0.144		spu:p	26.28	0.092
<i>Average</i>	31.58	0.257		<i>Average</i>	26.56	0.232

Two tailed, two sample *t*-tests on LND: $t(8) = 1.00, p > .34$, and CBP: $t(8) = 1.99, p > .08$

REFERENCES

- Alderete, John. 1997. Dissimilation as local conjunction. *Proceedings of the North Eastern Linguistic Society* 27.17–32.
- Baayen, R. Harold, Ricahrd Piepenbrock, and Leon Gulikers. 1995. *The CELEX Lexical Database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium.
- Baertsch, Karen, and Stuart Davis. 2003. The split margin approach to syllable structure. *ZAS Papers in Linguistics* 32.1–14.
- Bailey, Todd M., and Ulrike Hahn. 1998. Determinants of wordlikeness. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, ed. by Morton Ann Gernsbacher and Sharon J. Derry, 90–95. London: Lawrence Erlbaum Associates.
- Bailey, Todd M. and Ulrike Hahn. 2001. Determinants of wordlikeness: phonotactics or lexical neighborhoods. *Journal of Memory and Language* 44.568–591.
- Balota, David A., and James I. Chumbley. 1984. Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance* 10.340–357.
- Berent, Iris, Daniel L. Everett, and Joseph Shimron. 2001. Do phonological representations specify variables? Evidence from the Obligatory Contour Principle. *Cognitive Psychology* 42.1–60.
- Berent, Iris, Gary F. Marcus, Joseph Shimron, and Adamantios Gafos. 2002. The scope of linguistic generalizations: evidence from Hebrew word formation. *Cognition* 83.113–139.
- Berent, Iris, and Joseph Shimron. 1997. The representation of Hebrew words: evidence from the obligatory contour principle. *Cognition* 64.39–72.
- Berent, Iris, Joseph Shimron, and Vered Vaknin. 2001. Phonological constraints on reading: evidence from the Obligatory Contour Principle. *Journal of Memory and Language* 44.644–665.
- Berent, Iris, Donca Steriade, Tracy Lennertz, and Vered Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104.591–630.
- Berkley, Deborah M. 2000. *Gradient Obligatory Contour Principle effects*. Evanston: North Western University dissertation.
- Boersma, Paul. 1998. *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics
- Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32.45–86.
- Broselow, Ellen. 1991. *The structure of fricative-stop onsets*. Stony Brook: State University of New York, MS.
- Browne, Wayles. 1981. Slavic -ba and English *slil: two persistent constraints. *Folia Slavica* 4.219–225.
- Burzio, Luigi. 1995. The rise of Optimality Theory. *Glott International* 1(6).3–7.

- Bybee, Joan. 2006. From usage to grammar: the mind's response to repetition. *Language* 82.711–733.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.
- Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: The MIT Press.
- Chomsky, Noam and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.
- Clements, George N., and Samuel J. Keyser. 1983. *CV phonology: a generative theory of the syllable*. Cambridge, MA: The MIT Press.
- Coetzee, Andries W. 2004. *What it means to be a loser: non-optimal candidates in Optimality Theory*. Amherst: University of Massachusetts dissertation.
- Coetzee, Andries W. 2005. The OCP in the perception of English. *Prosodies*, ed. by Sonia Frota, Marina Vigario, and Maria Joao Freitas, 223–245. New York: Mouton de Gruyter.
- Coetzee, Andries W. 2006. Variation as accessing 'non-optimal' candidates. *Phonology* 23.337–385.
- Coetzee, Andries W. to appear. Grammar is both categorical and gradient. *Phonological Argumentation*, ed. by Stephen Parker. London: Equinox.
- Coetzee, Andries W. and Joe Pater. to appear. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory*.
- Coleman, John, and Janet Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. *Computational Phonology. Third Meeting of the ACL Special Interest Group in Computational Phonology*, 49–56. Somerset: Association for Computational Linguistics.
- Connine, Cynthia M., Debra Titone., and Jian Wang. 1993. Auditory word recognition: extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19.81–94.
- Cutler, Anne, Jacques Mehler, Dennis Norris, and Juan Seguí. 1987. Phoneme identification and the lexicon. *Cognitive Psychology* 19.141–177.
- Davidson, Lisa. 2006. Phonology, phonetics, or frequency: influences on the production of non-native sequences. *Journal of Phonetics* 34.04–137.
- Davis, Stuart. 1982. On the internal organization of syllable constituents. *Coyote Papers 3: Working Papers in Linguistics of the University of Arizona*, ed. by Thomas G. Larson, 3–22. Tucson: University of Arizona.
- Davis, Stuart. 1988a. Syllable onsets as a factor in stress rules. *Phonology* 5.1–19.
- Davis, Stuart. 1988b. *Topics in syllable geometry*. Outstanding Dissertation in Linguistics. New York: Garland.
- Davis, Stuart. 1989. Cross-vowel phonotactic constraints. *Computational Linguistics* 15.109–110.

- Davis, Stuart. 1991. Coronals and the phonotactics of nonadjacent consonants in English. *The special status of coronals: internal and external evidence*, ed. by Carole Paradis and Jean-Francois Prunet, 49–60. San Diego: Academic Press.
- de Lacy, Paul. 2007. Freedom, interpretability, and the loop. *Freedom of Analysis*, ed. by Patrik Bye, Martin Krämer, and Sylvia Blaho, 86–118. Amsterdam: John Benjamins.
- Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier, and Jacques Mehler. 1999. Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25.1568–1578.
- Dupoux, Emmanuel, Christophe Pallier, Kazuhiko Kakehi, and Jacques Mehler. 2001. New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes* 16.491–505
- Epstein, Samuel D. 1990. Differentiation and reduction in syntactic theory: a case study. *Natural Language and Linguistic Theory* 8.313–323. [Reprinted as Epstein 2000.]
- Epstein, Samuel D. 2000. Differentiation and reduction in syntactic theory: a case study. *Essays in syntactic theory*, ed by Samuel D. Epstein, 83–93. New York: Routledge. [Reprint of Epstein 1990.]
- Firth, John R. 1935. The use and distribution of certain English sounds. *English Studies: A Journal of English Letters and Philology* 17.8–18.
- Firth, John R. 1935–1937. Alphabets and phonology in India and Burma. *Bulletin of the School of Oriental Research* 8.517–546.
- Frisch, Stefan A., Janet B. Pierrehumbert, and Michael Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22.179–228.
- Frisch, Stefan A., Nathan R. Large, Bushra A. Zawaydeh, and David B. Pisoni. 2001. Emergent phonotactic generalizations in English and Arabic. *Frequency and the emergence of linguistic structure*, ed by Joan L. Bybee and Paul Hopper, 159–179. Amsterdam: John Benjamins.
- Frisch, Stefan A., and Bushra A. Zawaydeh. 2001. The psychological reality of OCP-Place in Arabic. *Language* 77.91–106.
- Fudge, Erik C. 1969. Syllables. *Journal of Linguistics* 5.253–286.
- Fujimura, Osamu, and Lovins, Julie B. 1978. Syllables as concatenative phonetic units. *Syllables and segments*, ed. by Alan Bell and Joan B. Hooper, 107–120. Amsterdam: North-Holland Publishing Company.
- Ganong, William F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6.110–125.
- Gouskova, Maria. 2001. Falling sonority onsets, loanwords, and syllable contact. *Papers from the 37th Meeting of the Chicago Linguistic Society. Vol. 1: The Main Session*, 175–185.
- Greenberg, Joseph H. 1978. Some generalizations concerning initial and final consonant clusters. *Universals of human language. Volume 2*, ed. by Edith A. Moravcsik, 243–279. Stanford: Stanford University Press.

- Halle, Morris. 1962. Phonology in a generative grammar. *Word* 18.54–72.
- Hay, Jennifer, Janet B. Pierrehumbert, and Mary E. Beckman. 2003. Speech perception, well-formedness and the statistics of the lexicon. *Phonetic interpretation: papers in laboratory phonology VI*, ed. by John Local, Richard Ogden, and Rosalind Temple, 58–74. Cambridge: Cambridge University Press.
- Hayes, Bruce. 1985. *A metrical theory of stress rules*. New York: Garland.
- Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: the early stages. *Fixing priorities: constraints in phonological acquisition*, ed. by René Kager, Joe Pater, and Wim Zonneveld, 158–203. Cambridge: Cambridge University Press.
- Hayes, Bruce, Robert M. Kirchner, and Donca Steriade (eds.) 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.
- Hayes, Bruce and Colin Wilson. to appear. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*.
- Hockett, Charles F. 1955. *A manual of phonology*. Baltimore: Waverly Press.
- Hume, Elizabeth V. and Keith Johnson (eds.) 2001. *The role of speech perception in phonology*. San Diego: Academic Press.
- Itô, Junko. 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* 7.217–259.
- Itô, Junko, and Mester, Armin. 1997. Sympathy Theory and German truncations. *University of Maryland Working Papers in Linguistics 5*, ed. by Viola Miglio and Bruce Moren, 117–139. College Park: University of Maryland.
- Itô, Junko, and Mester, Armin. 1998. *Markedness and word structure: OCP effects in Japanese*. Santa Cruz: University of California, Santa Cruz., MS
- Itô, Juno, and Armin Mester. 1999. The structure of the phonological lexicon. *The handbook of Japanese linguistics*, ed. by Natsuko Tsujimura, 62–100. Oxford: Blackwell.
- Kabak, Baris & William J. Idsardi 2007. Perceptual distortions in the adaptation of English consonant clusters: syllable structure or consonantal contact constraints? *Language and Speech* 50.23–52.
- Kager, René and Keren Shatzman. 2007. Phonological constraints in speech processing. Paper presented at workshop in *Experimental Approaches to Optimality Theory*, Ann Arbor.
- Kahn, Daniel. 1980. *Syllable-based generalizations in English phonology*. New York: Garland.
- Kawahara, Shigeto, Hajime Ono, and Kiyoshi Sudo. 2006. Consonant co-occurrence restrictions in Yamato Japanese. *Japanese/Korean Linguistics* 14.27–38.
- Kisseberth, Charles. 1970. On the functional unity of phonological rules. *Linguistic Inquiry* 1.291–306.
- Lamontagne, Greg. 1993. *Syllabification and consonant cooccurrence conditions*. Amherst: University of Massachusetts dissertation.

- Legendre, Geraldine, Antonella Sorace, and Paul Smolensky. 2006. The Optimality Theory-Harmonic Grammar connection. *The harmonic mind. Volume 2*, ed. by Geraldine Legendre and Paul Smolensky, 339–402. Cambridge, MA: The MIT Press.
- Luce, Paul A. 1986. *Neighborhoods of words in the mental lexicon*. Bloomington: Indiana University, Department of Psychology, MS.
- Luce, Paul A., and David B. Pisoni. 1998. Recognizing spoken words: the neighborhood activation model. *Ear and Hearing* 19.1–35.
- Massaro, Dominic W., and Michael M. Cohen 1983 Phonological context in speech perception. *Perception and Psychophysics* 34.338–348.
- McCarthy, John J. 1985. *Formal problems in Semitic phonology and morphology*. New York: Garland.
- McCarthy, John J. 1986. OCP effects: gemination and antigemination. *Linguistic Inquiry* 7:207–263.
- McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. *Papers in laboratory phonology III*, ed. by Patricia Keating, 191–283. Cambridge: Cambridge University Press.
- McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.
- McCarthy, John J. 2003. *Rotuman consonant cooccurrence restrictions*. Amherst: University of Massachusetts, MS.
- McCarthy, John J., and Alan Prince. 1993. *Prosodic morphology I: constraint interaction and satisfaction*. Report no. RuCCS-TR-3. New Brunswick: Rutgers University Center for Cognitive Science.
- McCarthy, John J. and Alan Prince. 1994. The emergence of the unmarked: optimality in prosodic morphology. *Proceedings of the North Eastern Linguistics Society* 24.333–379.
- McCarthy, John J. and Alan Prince. 2003. The emergence of the unmarked. *Optimality Theory in phonology: a reader*, ed. by John J. McCarthy, 483–494. Oxford: Blackwell.
- Moreton, Elliott. 2002. Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84.55–71.
- Moreton, Elliott. 2006. *Underphonologization and modularity bias*. Chapel Hill: University of North Carolina, MS.
- Myers, Scott. 1997. OCP effects in Optimality Theory. *Natural Language and Linguistic Theory* 15.847–892.
- Newman, Rochelle S., James R. Sawusch, and Paul A. Luce. 1997. Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology* 23.873–889.
- Odden, David. 1986. On the role of the Obligatory Contour Principle in phonological theory. *Language* 62.353–383.

- Ohala, John J. 1986. Consumer's guide to evidence in phonology. *Phonology Yearbook* 3:3–26.
- Padgett, Jaye. 1995. *Stricture in feature geometry*. Stanford: CSLI.
- Pater, Joe. 2005. Learning a stratified grammar. *Proceedings of the 29th Boston University Conference on Language Development*, ed. by Alejna Brugos, Manuella R. Clark-Cotton, and Seungwan Ha, 482–492. Somerville: Cascadilla Press.
- Pater, Joe and Andries W. Coetzee. 2005. Lexically specific constraints: gradience, learnability, and perception. *Proceedings of the Korea International Conference on Phonology*, 85–119. Seoul: The Phonology-Morphology Circle of Korea.
- Pater, Joe, Rajesh Bhatt and Christopher Potts. 2007. *Linguistic optimization*. Amherst: University of Massachusetts, MS.
- Peperkamp, Sharon, Rozenn Le Calvez, Jean-Pierre Nadal, and Emmanuel Dopoux. 2006. The acquisition of allophonic rules: statistical learning with linguistic constraints. *Cognition* 101.B31–B42.
- Pierrehumbert, Janet B. 1993. Dissimilarity in the Arabic verbal roots. *Proceedings of the North Eastern Linguistic Society* 23.367–381.
- Pierrehumbert, Janet B., Mary E. Beckman, and D. Robert Ladd. 2000. Conceptual foundations of phonology as a laboratory science. *Phonological knowledge: conceptual and empirical issues*, ed. by Noel Burton-Roberts, Philip Carr, and Gerard Docherty, 273–303. Oxford: Oxford University Press.
- Pierrehumbert, Janet B., Jennifer Hay, and Mary E. Beckman. 2004. Speech perception, well-formedness, and the statistics of the lexicon. *Papers in Laboratory Phonology VI*, ed. by John Local, Richard Ogden, and Rosalind Temple, 58–74. Cambridge: Cambridge University Press.
- Pitt, Mark.A. 1998. Phonological processes and the perception of phonologically illegal consonant clusters. *Perception and Psychophysics* 60.941–951.
- Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: constraint interaction in generative grammar*. Report no. RuCCS-TR-2. New Brunswick: Rutgers University Center for Cognitive Science. [Reprinted as Prince and Smolensky 2004.]
- Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: constraint interaction in generative grammar*. Oxford: Blackwell. [Reprint of Prince and Smolensky 1993.]
- Prince, Alan, and Bruce Tesar. 2004. Learning phonotactic distributions. *Fixing priorities: constraints in phonological acquisition*, ed. by In Rene Kager, Joe Pater, and Wim Zonneveld, 245–291. Cambridge: Cambridge University Press.
- Sagey, Elisabeth 1988. On the ill-formedness of crossing association lines. *Linguistic Inquiry* 19.109-118.
- Selkirk, Elisabeth O. 1982. *The syntax of words*. Cambridge, MA: The MIT Press.
- Silva, David J. 2004. Quantified optimality and the phonological parsing of Korean SOV sentences. *Korean Linguistics* 12.25–54.

- Smolensky, Paul. 1995. *On the internal structure of the constraint component CON of UG*. Talk presented at UCLA, April 7, 1995.
- Smolensky, Paul. 1996. On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27.720–731.
- Smolensky, Paul, and Géraldine Legendre. 2006. *The harmonic mind: from neural computation to Optimality-Theoretic grammar. Volume 1: cognitive architecture. Volume 2: linguistic and philosophical implications*. Cambridge, MA: The MIT Press.
- Spang-Hanssen, Henning. 1959. *Probability and structural classification in language description*. Copenhagen: Rosenkilde and Bagger.
- Tesar, Bruce. 1998. An iterative strategy for language learning. *Lingua* 104.131–145.
- Tesar, Bruce and Paul Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29.229–268.
- Tesar, Bruce, and Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, MA: The MIT Press.
- Treiman, Rebecca, Brett Kessler, Stephanie Knewasser, Ruth Tincoff, and Margo Bowman. 2000. English speakers' sensitivity to phonotactic patterns. *Papers in laboratory phonology V: acquisition and the lexicon*, ed. by Michael B. Broe and Janet B. Pierrehumbert, 269–282. Cambridge: Cambridge University Press.
- Twaddell, William F. 1939. Combinations of consonants in stressed syllables in German. *Acta Linguistica* 1.189–199.
- Twaddell, William F. 1940. Combinations of consonants in stressed syllables in German. *Acta Linguistics* 2.31–50.
- Uhlenbeck, Eugenius M. 1949. *De structuur van het Javaanse morpheem*. Bandung: Nix.
- Vitevitch, Michael S., and Paul A. Luce. 1998. When words compete: levels of processing in spoken word recognition. *Psychological Science* 9.325–329.
- Vitevitch, Michael S., and Paul A. Luce. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40.374–408.
- Wilson, Colin. 2006. Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* 30.945–982.
- Yip, Moira. 1989. Feature geometry and co-occurrence restrictions. *Phonology* 6.349–374.
- Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: corpus and survey evidence from Tagalog infixation. *Language* 83.277–316.