CHAPTER 6

WELL-FORMEDNESS JUDGMENTS AND LEXICAL DECISION

Language users have the ability to compare different linguistic constructs for their relative well-formedness. In the syntactic literature, for instance, there is a long tradition of marking sentences with different diacritics to indicate their relative well-formedness (*, ?, ??) (Epstein, 1990, 2000, Schütze, 1996). Although relative well-formedness has received less attention in phonology than in syntax, there have been some studies that address this issue for phonology (e.g. Berent and Shimron, 1997, Berent *et al*., 2001a, 2001b, 2002, Frisch *et al*., 2001, Frisch and Zawaydeh, 2001, Hayes, 1997, Hayes and MacEachern, 1996, 1997, Hayes, 1998, Pierrehumbert *et al*., In press).

A very salient property of these judgments is that they always reflect finer distinctions than the categorical distinction between the grammatical (possible) and ungrammatical (impossible). These judgments show a wide array of gradient well-formedness differences between forms. Linguistic theory in the generative tradition has focused strongly on the categorical grammatical/ungrammatical distinction, so that the gradient nature of these judgments has resulted in them receiving relatively little attention in linguistic theory. This is unfortunate, since these judgments are usually strongly influenced by grammar. As such they provide a rich and still largely untapped source of information about the linguistic competence of language users.

The speed with which language users identify non-words in lexical decision tasks is another aspect of linguistic performance that is non-categorical in nature. The speed of reaction is measured in terms of reaction time, and time is a continuous variable. This measure is therefore necessarily non-categorical.

Lexical decision reaction times have long been used in psycholinguistic studies as a way in which to probe into the workings of the speech processing module (e.g. Balota and Chumbley, 1984, Berent and Shimron, 1997, Berent *et al*., 2001b, Shulman and Davison, 1977, Stone and Van Orden, 1993, Vitevitch and Luce, 1999) . However, again since reaction times are strictly non-categorical while linguistic theories in the generative tradition are categorical, linguistic theory has not paid much attention to lexical decision reaction times as a possible source of information about linguistic competence.

In this chapter of the dissertation I will argue that both well-formedness judgments and reaction times in lexical decision tasks are valuable sources of information about linguistic competence. I will also argue that it is possible to account for these non-categorical phenomena within the rank-ordering model of EVAL.

About well-formedness judgments I will argue that the grammar compares the different non-words in terms of their markedness and imposes a rank-ordering on the non-words. Non-words that receive a higher well-formedness rating are then simply tokens that occupy a relatively high slot in the rank-ordering.

With regard to lexical decision I will argue that language users use the well-formedness of a token as one of the considerations when making lexical decisions. The less well-formed a non-word token is in terms of the grammar of the language, the less seriously the language user will entertain the possibility that the token could be a word, and the quicker the token will be rejected as a non-word. In the rank-ordering model of EVAL, this can be explained as follows: The grammar compares the non-word tokens and imposes a rank-ordering on them. The lower slot a non-word occupies on this rank-ordering, the less well-formed it is, and the quicker it will be rejected as a non-word.

The rest of this chapter is structured as follows: In §1 I consider some theoretical preliminaries about how the rank-ordering model of EVAL can account for data reported in this chapter. Sections §2 and §3 form the central part of this chapter. In §2 I illustrate how the rank-ordering model of EVAL can account for the relative well-formedness judgments and lexical decision reaction times associated with the Obligatory Contour Principle (OCP) (Goldsmith, 1976, Leben, 1973, McCarthy, 1986, Yip, 1988) in Hebrew. In this section I rely on the data reported by Berent *et al.* (2001a, 2001b, 2002, Berent and Shimron, 1997). In §3 I report on a set of experiments in which I replicated Berent *et al.*'s results for a different phonotactic constraint in English. In §4 alternative accounts of the data are considered. Section §5 then contains a summary and conclusion.

1. Preliminary considerations

This section of the chapter is structured as follows: In §1.1 I give an illustration of how the rank-ordering model of EVAL can be used to account for well-formedness judgments and reaction times in lexical decision tasks. In §1.2 I then consider the relationship between lexical statistics and the grammar. There is ample evidence that statistical patterns in the lexicon influence both well-formedness judgments and lexical decision reaction times. Section §1.2 argues that grammar, in addition to these lexical statistics, also contributes towards determining well-formedness judgments and lexical decision reaction times.

1.1 Well-formedness and lexical decision in a rank-ordering model of EVAL

I propose two extensions to the classic OT grammar in this dissertation (see Chapter 1 §1). First, I argue that EVAL can compare not only candidate output forms generated for some input (generated comparison set), but that EVAL can compare any set of forms – even forms that are not related to each other via a shared input (non-generated comparison set). Secondly, I argue that EVAL imposes a harmonic rank-ordering on the full candidate set. EVAL therefore does more than to distinguish the best candidate from the rest. EVAL also imposes a rank-ordering on the non-best candidates. In this section I show how these two extensions to a classic OT grammar enable us to account for well-formedness judgments and lexical decision reaction times. I will use a made-up example for this purpose.

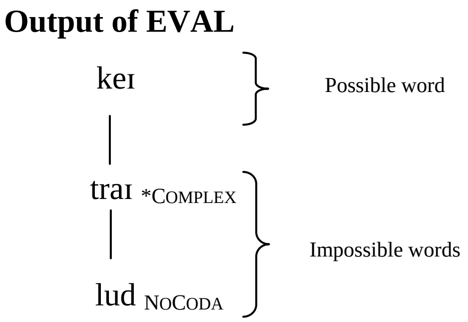
Consider a language L_1 with the ranking ||NOCODA o *COMPLEX o DEP||. In L_1 neither codas nor complex onsets are allowed – both are avoided by epenthesis. Assume that none of the forms [keɪ], [traɪ] or [lud] is an actual word of L_1. Of these three forms only [keɪ] is a possible word of L_1. [traɪ] is not a possible word because it has a complex margin, and [lud] is not a possible word because it has a coda.

Now consider a language L_2 with the ranking ||NOCODA o DEP o *COMPLEX||. Like L_1, L_2 does not allow codas – codas are avoided by epenthesis. However, unlike L_1, L_2 does tolerate complex onsets – because DEP outranks *COMPLEX. Assume that none of the forms [keɪ], [traɪ] or [lud] is an actual word of L_2. Of these three [lud] is not even a possible word, since it has a coda. However, the other two are both possible words.

In (1) these three non-words are compared with each other in each of L_1 and L_2. Next to the tableaux is a graphic representation of the rank-ordering that EVAL will impose on these three candidates in each language. (On the typographical conventions used in these tableaux, see Chapter 1 §2.2.)

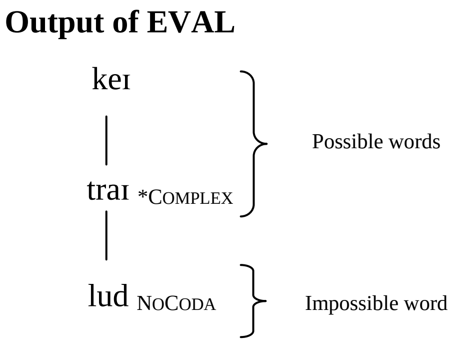
(1) a. **Comparison in *L*$_1$**

		NoCoda	*Complex	Dep
1	keɪ			
2	traɪ		*	
3	lud	*		



b. **Comparison in *L*$_2$**

		NoCoda	Dep	*Complex
1	keɪ			
2	traɪ			*
3	lud	*		



In both *L*$_1$ and *L*$_2$ EVAL imposes the following rank-ordering on these three forms: |keɪ ™ traɪ ™ lud|. If language users had to rate these three forms as to their relative well-formedness, the prediction is that [keɪ] would be rated best, then [traɪ], and finally [lud]. The prediction is therefore that language users will make more than a categorical distinction between possible and impossible words. The *L*$_1$-example shows that in addition to the distinction between possible ([keɪ]) and impossible ([traɪ] and [lud]), language users are predicted to also distinguish between the two impossible words in terms of their well-formedness. The *L*$_2$-example shows that in addition to the distinction between possible ([keɪ] and [traɪ]) and impossible ([lud]), language users are predicted to also distinguish between the two possible words in terms of their well-formedness.

What about a lexical decision task? The proposal is that language users employ, among other things, the information provided by grammar when making lexical decisions. The more well-formed a token is, the more seriously the language users will consider the possibility that the token might be a word. In general, it is therefore expected that a non-word that is more well-formed will be rejected more slowly than a non-word that is less well-formed. In both L_1 and L_2 the prediction is that [keɪ] will have the slowest rejection time, with [traɪ] faster, and [lud] the fastest. Again, we predict that language users will not only make the categorical distinction between possible and impossible words. In L_1 they will treat different impossible words differently based on their relative well-formedness – although both [traɪ] and [lud] are impossible as words of this language, [lud] is less well-formed and will therefore be rejected faster than [traɪ]. In L_2 they will treat different possible words differently based on their relative well-formedness – although both [keɪ] and [traɪ] are possible as words of this language, [traɪ] is less well-formed and will therefore be rejected more quickly than [keɪ].

In this explanation we use both of the extensions to a classic OT grammar that are argued for in this dissertation. First, about the comparison sets: The three forms compared in (1) are not related to each other via a shared input. If EVAL could only compare candidate output forms generated by GEN for some input, then the comparison in (1) would not even be possible.

Had EVAL only distinguished the best candidate from the rest of the candidates in a comparison set, then the tableaux in (1) would only have given us evidence that [keɪ] is more well-formed than both [traɪ] and [lud]. But we would not have known anything about how [traɪ] and [lud] are related to each other. In L_1 we would then expect language users to treat the two impossible words the same. And in L_2 we would expect language users to treat the possible word [traɪ] the same as the impossible word [lud]. In both languages the relative well-formedness difference between the two forms [traɪ] and [lud] would not be predicted at all.

In this chapter I will discuss two examples (one from Hebrew and one from English) showing that (i) language users can compare morphologically unrelated forms for their well-formedness, and that (ii) these comparisons are done across more than just two levels. In both examples we find evidence that language users distinguish possible and impossible words in terms of their well-formedness. However, the Hebrew example shows that language users also make well-formedness distinctions within the set of possible words, and the English example shows that language users make well-formedness distinctions in the set of impossible words. This serves as evidence for the two extensions to the classic OT grammar argued for in the dissertation.

1.2 Factors in the processing of non-words

One claim that I will make in this chapter is that language users employ the grammar of their language in the processing of non-words. The response patterns in the gradient well-formedness and lexical decision experiments are therefore predicted to reflect grammatical effects. The claim that grammar influences phonological processing is not novel. For more claims that phonological processing is mediated via grammar see Berent *et al*. (2001a, 2001b, 2002, Berent and Shimron, 1997), Brown and Hildum (1956), Coetzee (to appear), Frisch and Zawaydeh (2001), Moreton (2002a, 2002b, 2003), etc. For claims that grammar plays a much smaller role in phonological processing see Coleman and Pierrehumbet (1997), Luce (1986), Newman *et al*. (1997), Pierrehumbert *et al*. (In press), Treiman *et al*. (2000, Treiman, 1983,), Vitevitch and Luce (1998, 1999), etc.

The factors that have been claimed to influence phonological processing can be classified broadly into two categories, namely lexical and grammatical. In terms of lexical influences, it has been shown that statistical patterns extracted over the lexicon influence phonological processing. In particular, two kinds of lexical statistics that influence phonological processing have been identified, namely *lexical neighborhood density* and *transitional probabilities*. Rather than reviewing the large body of literature about the influence of these two kinds of lexical statistics, I will discuss a few representative examples from the literature.

1.2.1 Lexical statistics

Lexical neighborhood density. The lexical neighborhood density of a token is a function of both the number of words similar to the token and the usage frequency of these words. The more words similar to some form and the more frequent these words are, the denser the lexical neighborhood of this form will be.[1] Both words and non-words have lexical neighborhoods – the lexical neighborhood of a token (word or non-word) is comprised of all those words in the lexicon that are similar to the token to some specified degree. Lexical neighborhood density has been shown to influence the phonological processing of non-words in several ways. Luce and Pisoni (1998), for instance, performed a speeded lexical decision experiment in which they presented their subjects with non-words that differed in lexical neighborhood densities. They found that (i) non-words that occupy

---

[1] See Luce (1986) and Luce and Pisoni (1998) for an example of exactly how lexical neighborhood density can be calculated. See also §3.2 below.

denser lexical neighborhoods were consistently responded to more slowly than non-words that inhabit sparser lexical neighborhoods, and (ii) non-words that occupy a denser neighborhood were consistently responded to less accurately than non-words from sparser neighborhoods. They interpret these results as follows: The nature of the lexical decision task requires that the percept be checked against entries in the mental lexicon.[2] However, listeners do not check the percept against all entries in the lexicon, but only against entries that are similar to the percept to some specified degree (i.e. against words in the immediate lexical neighborhood of the token). Non-word tokens that inhabit denser lexical neighborhoods therefore need to be checked against more lexical entries than non-words that inhabit sparser lexical neighborhoods, with the result that non-words from sparse lexical neighborhoods can be rejected more quickly than non-words from dense lexical neighborhoods. Similarly, the larger number of lexical entries against which non-words from denser neighborhoods need to be checked results in more opportunities for mistakenly identifying the non-word percept as an actual lexical entry. For more studies that manipulated the lexical neighborhood density of tokens in lexical decision tasks, see *inter alia* Balota and Chumbley (1984), and Vitevitch and Luce (1999).

The results of these lexical decision experiments suggest that denser lexical neighborhoods inhibit the processing of non-words – denser neighborhoods result in slower and less accurate processing. However, several studies have also been reported that seem to present evidence to the contrary – namely that a dense lexical neighborhood

---

[2] This assumption of Luce and Pisoni (that lexical decision by its very nature requires lexical access) is not a necessary assumption. If a percept violates a phonotactic or prosodic constraint that all words in the language must obey, then it can be rejected as a non-word without accessing the lexicon. All of the tokens used by Luce and Pisoni were possible words of English. Their assumption that lexical access is necessary for lexical decision is therefore true for their experimental design. See also Balota and Chumbley (1984).

aid in the processing of non-words. It has, for instance, been shown (i) that non-words from denser lexical neighborhoods are judged more word-like than non-words from sparser neighborhoods (Bailey and Hahn, 1998), and (ii) that non-words from denser neighborhoods are shadowed (repeated) significantly faster than non-words from sparser neighborhoods (Vitevitch and Luce, 1998). It seems to be that a dense lexical neighborhood sometimes aids in the processing of non-words, and sometimes inhibits the processing of non-words. Vitevitch and Luce (1999) argue for a principled distinction between the tasks in which lexical neighborhood density aids processing and the tasks in which lexical neighborhood density inhibits processing. They claim that the inhibitory effect of lexical neighborhood density is observed only in tasks that require lexical access. Lexical decision requires lexical access,[3] and in lexical decision tasks a denser neighborhood counts against a token. However, word-likeness judgments and shadowing do not require lexical access, and in these tasks a denser lexical neighborhood seems to count in favor of a token.

Transitional probability. Based on these results, Vitevitch and Luce (1999) and Newman *et al*. (1997) claim that there are two kinds of lexical statistics that can influence processing of non-words, namely lexical neighborhood density and transitional probabilities. Transitional probability refers to the likelihood of two sounds occurring next to each other in a specific order – i.e. we can compute the probability of, for instance, the sound [k] occurring before the sound [æ] in English. Since tokens with many and frequent neighbors usually also consist of frequent sounds in frequent combinations, there is generally a high positive correlation between the lexical neighborhood density and the

---

[3] But see the previous footnote.

transitional probabilities of a token – a token with a higher lexical neighborhood density is very likely also to have high transitional probabilities.

Newman *et al.* (1997) and Vitevitch and Luce (1999) claim that information about the transitional probabilities can be accessed without lexical access. If the lexicon is not accessed (as is the case in word-likeness judgments and shadowing tasks), then the lexical neighbors of a token are not activated. A token with many neighbors and a token with few neighbors are therefore equal on this front – neither has to compete with lexical neighbors. In these tasks higher transitional probabilities then aid in the processing of a token. The advantage of tokens with high lexical neighborhoods in these tasks results not from their lexical neighborhoods, but from their transitional probabilities which correlate with their lexical neighborhood density.

Transitional probabilities are calculated in two different ways in the literature. One method is to calculate these probabilities simply over the entries in the lexicon. In this kind of calculation the frequency with which different lexical tokens are used in the language is not taken into consideration in the calculation. To continue with the example used above, in calculating the probability of [k] preceding [æ], a relatively scarce word such as *catamaran* will count equally as much as a relatively frequent word such as *can*. Transitional probabilities are calculated according to this method by, for instance, Frisch and Zawaydeh (2001) for Arabic and Berent *et al*. (2001a) for Hebrew. Another way in which transitional probabilities are calculated, is by weighting the contribution of individual tokens from the lexicon according to their frequency of usage. A more frequent word such as *can* will then contribute more than a less frequent word such as *catamaran*. This method was used for English, *inter alia*, by Vitevitch and Luce (1999), Pierrehumbert *et al*. (In press), and Treiman *et al*. (2000). Bailey and Hahn (1998:93) calculated transitional probabilities for English according to both methods, and found a very high correlation between the results of the two calculations (r^2 = 0.96). Although it seems best to take token frequency into consideration, this result of Bailey and Hahn shows that studies that did not take token frequency into consideration are most likely very comparable to studies that did.[4]

Transitional probabilities have been shown to contribute to phonological processing of non-words in several different ways. Vitevitch and Luce (1998), for instance, performed a speeded repetition task in which they presented their subjects with non-words with high transitional probabilities and non-words with low transitional probabilities. They found that non-words with high transitional probabilities were repeated significantly faster than non-words with low transitional probabilities. Their results therefore show that higher transitional probabilities aid in the processing of non-words.

Transitional probability has also been shown to influence word-likeness or gradient well-formedness judgments. There are many studies that report this result – a few of the more important are Bailey and Hahn (1998), Coleman and Pierrehumbert (1997), Frisch *et al.* (2001), Pierrehumbert *et al.* (In press), Treiman *et al*. (2000). The basic result reported in all of these studies is that a non-word that has higher phoneme

---

[4] For some languages it is not practically possible to take token frequency into consideration. In order to do to this, information about the usage frequency of different words is required, and this kind of information is for most languages not available. For instance, it is not available for Hebrew or for Arabic. Frisch and Zawaydeh (2001) and Berent *et al*. (2001a) could therefore not take token frequency into consideration.

transitional probabilities is in general judged as more well-formed or more word-like than a non-word with lower transitional probabilities.

There is therefore evidence for the fact that both lexical neighborhood density and transitional probabilities influence the processing of non-words. However, there is also evidence (i) that the influence of these factors are typically rather small (they account for only a small fraction of the variation in the observed data), and (ii) that the influence of lexical statistics is not very robust and can easily be reduced or even completely eliminated by varying the experimental task conditions. In what follows I will briefly present evidence for these two claims before moving on to discussing the evidence for the influence of grammar on the processing of non-words.

Bailey and Hahn (1998) conducted a word-likeness experiment with the express purpose of determining how much transitional probabilities contributed towards explaining the variation in the response data. They collected relative well-formedness judgments on 291 possible non-words of English. They then calculated the weighted phoneme transitional probabilities of the non-words, and performed a regression analysis on the response data with the transitional probabilities of the tokens as the independent variable. They found that transitional probability was significantly correlated with well-formedness judgments. However, they also found that transitional probabilities explained only a very small portion of the variation in their response data ($r^2 = 0.09$, $p < 0.0001$). Although transitional probabilities do contribute towards determining well-formedness judgments, it is clear that they do not contribute much. There must be other factors that also contribute towards the results.

The contribution of transitional probability towards non-word processing is therefore small. In addition to it being small, it is also not a very robust effect. The effect of lexical statistics (both transitional probabilities and lexical neighborhood density) has been shown to decrease or even disappear completely in certain experimental designs. There are at least two situations in which the effects of lexical statistics decrease or disappear. The first is when the experimental setup is such that subjects do not dedicate enough processing resources to the processing of the non-word tokens. Pitt and Samuel (1993), for instance, found that adding a distractor task results in a reduction of the effects caused by lexical statistics. They conducted phoneme identification experiments and found that lexical statistics did influence perception.[5] However, when they added a distractor task (by requiring their listeners to perform rudimentary mathematical calculations during the phoneme identification task), the effect of lexical statistics was significantly reduced.

The second situation in which the effect of lexical statistics can disappear is when the stimulus sets used in experiments are not varied enough. Cutler *et al*. (1987) also conducted phoneme identification experiments. When they used stimulus sets that contain forms of varied prosodic shapes (monosyllabic, disyllabic, initial stress, final stress, etc.), they found that the response patterns were influenced by the lexical statistics.

---

[5] They presented their listeners with tokens ambiguous between two percepts. The one endpoint was a word and the other a non-word. The word endpoint is favored by a lexical bias (Ganong, 1980). Their listeners responded according to this bias – i.e. they were more likely to identify the phoneme such that a real word resulted. However, the advantage of the word endpoint was significantly reduced by the addition of the distractor task.

However, when the stimulus set contained tokens with only one prosodic shape (only monosyllables), the effect of lexical statistics disappeared.[6]

1.2.2 Grammar

Grammar has also been shown to influence the processing of non-words. The general idea behind this claim is that non-words are processed differently depending on how well they conform to grammar. There is evidence that non-words that do conform to the grammar (possible words) are processed differently than non-words that do not conform to the grammar (impossible words). But there is also evidence that grammar makes finer distinctions than simply between possible and impossible words. Even within the set of possible words, a distinction is made between non-words that are more well-formed and non-words that are less well-formed. The same is true for the set of impossible words – also here a distinction between more and less well-formed non-words arises. This has been shown to be true for gradient well-formedness judgments and lexical decision tasks. In the rest of this section I will discuss representative examples from the literature on how grammar influences phonological processing.

Frisch and Zawaydeh (2001) conducted a word-likeness experiment in which they found strong evidence that the grammar (in addition to lexical statistics) contributes towards how word-like language users consider a non-word to be. They investigated the sensitivity of Arabic speakers to the restrictions on the structure of verbal roots. Arabic verbal morphology, like that of other Semitic languages, is based on tri-consonantal verbal roots. The verbal roots are subject to a restriction that prohibits homorganic

---

[6] There are more experimental results reported in the literature that point to the fact that the contribution of lexical statistics to phonological processing is heavily task dependent (Eimas *et al.*, 1990, Eimas and Nygaard, 1992, Frauenfelder and Seguí, 1989).

consonants from occurring in the same root (because of some kind of OCP-effect).[7] Frisch and Zawaydeh chose non-words that violate the OCP and non-words that do not violate the OCP. The tokens were chosen such that forms that do violate the OCP had high transitional probabilities and dense lexical neighborhoods, and *vice versa* for forms that obey the OCP. Based on results such as that of Bailey and Hahn (1998) discussed just above, it is therefore expected that the OCP-violating tokens with the higher transitional probabilities and lexical neighborhood densities should be judged as more word-like than the non-violators. However, they found the opposite. The OCP-violating tokens were consistently judged as less word-like in spite of the fact that these tokens were favored by the lexical statistics. This shows (i) that grammar does contribute towards the processing of non-words, and (ii) that the effects of lexical statistics can be overridden by grammar – when there is a conflict between grammar and lexical statistics, grammar wins out.[8]

Berent *et al*. (2001a, 2001b, 2002, Berent and Shimron, 1997) conducted a series of experiments in which they report very similar findings for Hebrew.[9] Like Arabic, Hebrew verbal morphology is based on tri-consonantal roots. There are restrictions on the distribution of identical consonants in the tri-consonantal roots. Forms without identical consonants are possible words of Hebrew (i.e. [QiSeM] is a possible word). Also, forms with identical consonants in the positions usually occupied by the final two root

---

[7] The restriction is actually much more complicated than this. See Frisch *et al*. (2004), Greenberg (1950), McCarthy (1986, 1994), Pierrehumbert (1993) for a discussion of the details of the consonant co-occurrence restrictions in Arabic.

[8] See also Berent and Shimron (1997:55-56), Frisch *et al*. (2001:170, 174) and Pierrehumbert (1994:181) about the idea that the effects of grammar can override that of lexical statistics.

[9] The Berent *et al*. studies are discussed in detail in §2.2 below.

consonants are possible words of Hebrew (i.e. [QiSeS] is a possible word). However, forms with identical consonants in the positions usually occupied by the first two root consonants are not possible words of Hebrew (i.e. [QiQeS] is not a possible word).[10] Berent *et al*. conducted a series of experiments to test whether speakers of Modern Hebrew are sensitive to this restriction on possible words. They presented the subjects in their experiments with a stimulus list of non-words, some of which had no identical consonants, some with final identical consonants, and some with initial identical consonants. In the selection of stimuli, they controlled for the possible influence of lexical statistics. The task of the subjects in their experiments was to rate these non-words as to their word-likeness or well-formedness. Berent *et al*. found that possible words (no identical consonants and final identical consonants) were judged better than impossible words (initial identical consonants). However, they also found evidence for a distinction within the set of possible words. Forms with no identical consonants were judged better than forms with final identical consonants – even though both of these forms are possible words of Hebrew. They interpret this result as showing that even forms with final identical consonants violate some well-formedness constraints, and that speakers of Hebrew have access to this information. Grammar therefore makes finer distinctions than simply between possible and impossible words. Grammar also distinguishes within the set of possible words between more and less well-formed tokens.

Berent *et al*. (2001b) have also shown that grammar influences lexical decision. In a lexical decision experiment they presented their subjects with a list of tokens that

---

[10] See Gafos (1998, 2003), McCarty (1986) and Ussishkin (1999) for analyses of this phenomenon. See also §2.1 below for an alternative OT account of this restriction.

contained both words and non-words. They controlled for the possible influence of lexical statistics in the selection of the stimuli. The task of the subjects was to distinguish between words and non-words. Reaction times were measured. Berent *et al*. found evidence for a distinction between possible and impossible words. Impossible non-words (with identical initial consonants) were rejected significantly faster than possible non-words (with identical final consonants). The subjects in their experiment therefore used the information supplied by grammar when making lexical decisions. The less well-formed (actually ill-formed) tokens were considered less seriously as potential words, and therefore rejected more quickly. See also Stone and Van Orden (1993) for results that show that grammar influences lexical decision.

In summary we can say that both lexical statistics and grammar contribute towards the processing of non-words. Therefore, I do not claim that lexical statistics are irrelevant in the processing of non-words, but rather that lexical statistics alone are not sufficient to account for how non-words are processed. We also need to take the contribution of grammar into consideration.

In the next two sections of this chapter I discuss two examples that show that grammar influences the processing of non-words. In §2 I discuss the experiments of Berent *et al*. referred to above in more detail. These experiments show that: (i) Hebrew speakers employ grammar in the processing of non-words. (ii) In particular, these results show that Hebrew speakers make multi-level distinctions in terms of well-formedness. They not only distinguish between possible and impossible words, but that they make distinctions within the set of possible words. Possible words that are more well-formed formed according to the grammar are treated differently from possible words that are less well-formed. In §3 I then discuss a series of experiments that I conducted in which I replicated the results of Berent *et al*. with English speakers. These experiments show that: (i) English speakers use the information provided by grammar in the processing of non-words. (ii) Specifically, English speakers also make multi-level distinctions. In addition to distinguishing possible and impossible words, they also distinguish between different impossible words in terms of their well-formedness. Impossible words that are more well-formed are reacted to differently than impossible non-words that are less well-formed.

2. The OCP in the processing of non-words in Hebrew

One of the most striking features of Semitic morphology is that the overwhelming majority of verbal roots are tri-consonantal. In addition to this, there are strict limitations on the distribution of consonants between the three consonantal positions in the root (Frisch *et al*., 2004, Gafos, 2003, Greenberg, 1950, McCarthy, 1986, 1994, Morgenbrod and Serifi, 1981, Pierrehumbert, 1993). I will focus here on one of these distributional restrictions, namely on the restriction on the distribution of identical consonants in the root. Forms with identical consonants in the positions usually occupied by the first two root consonants are generally not allowed – i.e. *[QiQeS] is ill-formed. On the other hand, forms with identical consonants in the positions usually occupied by the last two root consonants are well-formed – i.e. [QiSeS] is acceptable. I will refer to forms such as [QiQeS] as initial-geminates, and to forms such as [QiSeS] as final-geminates. (This follows in the tradition of the classical grammars of Hebrew and Arabic in which these forms are called “geminates”.) In this section of the chapter, I will discuss experimental evidence showing that this restriction has psychological reality for speakers of Modern Hebrew. In particular, it determines how speakers of Modern Hebrew process non-words.

This section of the chapter is structured as follows: In §2.1 I develop an OT analysis of this asymmetrical distribution of identical consonants. The analysis will establish that: (i) forms without identical consonants (neither in initial nor in final consonantal slots), are the most well-formed; (ii) forms with identical consonants in the final two consonantal slots are less well-formed; (iii) and forms with identical consonants in the initial consonantal slots are the least well-formed – i.e. |QiSeM ™ QiSeS ™ QiQeS|. In §2.2 I then discuss the results of a series of experiments performed by Berent *et al*. (2001a, 2001b, 2002, Berent and Shimron, 1997) that confirm this harmonic rank-ordering.

2.1 Restrictions on identity in roots

The analysis that I present here follows broadly in the tradition of the OCP-based account suggested by McCarthy (1979, 1981, 1986). The basic assumptions that I make are that (i) geminate forms are derived from bi-consonantal roots, (ii) the final root consonant of bi-consonantal roots spreads in order to create a tri-consonantal stem; (iii) identical contiguous consonants are not allowed in Hebrew roots.

2.1.1 Bi-consonantal roots

McCarthy (1979, 1981, 1986) argued that geminate forms are derived from bi-consonantal roots via autosegmental spreading of the final consonant. Although there has not been agreement about the mechanism involved in copying/spreading of the final root consonant, the idea that these forms are derived from bi-consonantal roots has been widely accepted in the literature – see for instance Bat-El (1994), Gafos (1998, 2003),[11] and Ussishkin (1999). In the rest of this section I will present evidence in favor of the bi-consonantal root approach. The evidence in favor of the bi-consonantal root comes from two sources – from an active process of denominal verbal formation in Modern Hebrew, and from the results of psycholinguistic experiments with speakers of Modern Hebrew.

Modern Hebrew has an active morphological process of deriving verbs from nominals. Hebrew has many nominals with only two consonants, and verbs are also regularly created from these nominals. This shows that Hebrew grammar should allow for the possibility of creating verbs from bi-consonantal roots. The examples below are from Ussishkin (1999:405). See also Bat-El (1994) for a discussion of this word formation process.

(2) **Verbs derived from bi-consonantal nominals in Modern Hebrew**

Noun/Adjective		**Verb**	
sam	‘drug’	*simem*	‘to drug, to poison’
dam	‘blood’	*dimem*	‘to bleed’
mana	‘portion’	*minen*	‘to apportion’
t^{s}ad	‘side’	*t^{s}ided*	‘to side’
xad	‘sharp’	*xided*	‘to sharpen’

There is also evidence from psycholinguistic experiments that speakers of Modern Hebrew can form verbs from bi-consonantal roots. Berent *et al*. (2001a) presented speakers of Modern Hebrew with a fully conjugated verbal exemplar and a bare

---

[11] Gafos (2003) differs from the other sources cited in that he assumes that the second consonant in the bi-consonantal root is underlyingly specified as geminate (or bi-moraic).

consonantal root (neither the exemplar nor the consonantal root corresponded to actual Hebrew words). The task of the subjects was to conjugate the consonantal root in analogy to the fully conjugated exemplar. Half of the consonantal roots were tri-consonantal and half were bi-consonantal – 48 roots of each kind. Their subjects did not report any particular difficulty in conjugating the bi-consonantal roots, and the responses were in fact quite accurate – more than 93% of the responses were correct. The majority of the correct responses involved repetition of the second consonant.[12]

The denominal verbal derivation and the results of the Berent *et al*. study show that speakers of Modern Hebrew can derive final-geminates from bi-consonantal roots. The grammar of Hebrew should therefore allow for this possibility.

The derivation of geminate verbs from bi-consonantal roots is a necessary part of Hebrew grammar. Therefore, the account offered for the asymmetry in the distribution of identical consonants has to explain why it is always the second consonant that spreads. Abstracting away from the origin of the vowels, the account has to explain why /Q-S/ → [QiSeS] is observed but not /Q-S/ → [QiQeS].

Even though I will argue that geminate verbs are derived from bi-consonantal roots, this does not exclude the possibility of there being roots with identical contiguous

---

[12] Berent *et al*. counted four kinds of responses as correct. (i) Final gemination. (ii) Initial gemination. (iii) No gemination – there are some verbal forms in Hebrew that can be formed with only two consonants – especially when one of the root consonants is “weak” (typically a glide). (iv) Addition – when the subjects added a third consonant that was different from any of the two consonants in the root with which they were presented. When subjects used the addition option, they nearly always added a /y/. A /y/ in a Hebrew root is a “weak” consonant and is often subject to deletion. The responses were distributed as follows between these four categories (as a percentage of the total responses, both correct and incorrect):

Final gemination:	47%	No gemination:	32%
Initial gemination:	< 1%	Addition:	14%

consonants. Richness of the base requires that the grammar be able to derive phonotactically legal output forms from any input (Prince and Smolensky, 1993, Smolensky, 1996). The grammar should therefore also be able to handle input roots such as /Q-Q-S/ and /Q-S-S/. I will argue that, were roots such as these to exist, they will be unfaithfully mapped onto the surface. In particular, one of the identical consonants will be deleted so that the surface correspondent of the root has only two non-identical consonants. The second of the two remaining consonants of the root will then spread, just like in a bi-consonantal root. The result will be that bi-consonantal roots and tri-consonantal roots with identical initial or final consonants all map onto the same surface structure – i.e. /Q-S/, /Q-Q-S/ and /Q-S-S/ all map onto [QiSeS].

When presented with a final-geminate form such as [QiSeS], which of the three possible underlying representations (/Q-S/, /Q-Q-S/ or /Q-S-S/) will a Hebrew speaker assume? I will argue that language users rely on the mechanism of lexicon optimization (Prince and Smolensky, 1993, Smolensky, 1996) in selecting underlying representations. This mechanism will consider different possible underlying representations and select the one that results in the most harmonic mapping from underlying representation to surface form. I will the show that of the three possible underlying representations, the bi-consonantal /Q-S/ results in the most harmonic mapping to the surface form [QiSeS]. Hebrew speakers will therefore assume that the root of any surface geminate is a bi-consonantal form. Although the grammar can handle geminate roots such as /Q-Q-S/ and /Q-S-S/, the Hebrew lexicon does not actually contain such roots.

In the next section (§2.1.2) I will develop an explanation for why bi-consonantal roots are mapped onto tri-consonantal surface forms, putting aside until later (§2.1.3) the question of why geminate forms are derived from bi-consonantal rather than tri-consonantal roots.

2.1.2 From bi-consonantal roots to tri-consonantal stems

The discussion in the previous section shows that Hebrew grammar has to allow for the possibility that tri-consonantal verbal forms can be derived from bi-consonantal roots. Once this is accepted, we have two questions that need to be answered: (i) Why do the bi-consonantal roots not map onto bi-consonantal output forms? (ii) Why is it never the first consonant of the root that is doubled? I consider each of these questions in turn below.

Verbal stems in Semitic are generally required to end on a consonant (Gafos, 1998, McCarthy and Prince, 1990b).[13] This requirement is expressed by the constraint FINAL-C.[14] In order to see how this constraint can force doubling of the final consonant of a bi-consonantal root, consider the following example: One of the conjugations of

---

[13] This requirement is not that the word should end in a consonant, but the stem. Stem should here be interpreted as that form to which prefixes and suffixes attach. There are two reasons why this constraint cannot refer to the word: (i) Bi-consonantal roots show doubling of the final root consonant even when there are suffixes added to the stem, cf. [maS.Mi.Mim] (Berent *et al.*, 2001a:26). The doubling of the final /M/ in this form cannot be ascribed to a need for the word to end in a consonant. (ii) There are many words in Hebrew that end in vowels.

[14] The constraint FREE-V used by Prince and Smolensky in their analysis of Lardil truncation (Prince and Smolensky 1993: Chapter 7, no. (152)) can also be interpreted as a ban on (nominative noun) stems ending in vowels.

Ussishkin (1999:415) argues against FINAL-C on the grounds of it being a “templatic” constraint. He argues that it should be replaced by a constraint that he calls STRONG-ANCHOR-R. This constraint requires the rightmost consonant of the input to have a correspondent at the right edge of the stem. A mapping /Q-S, i-e/ → [Qi.Se] violates this constraint, since the rightmost consonant of the input /S/ does not have a correspondent at the right edge of the stem. This constraint is stated within the schema of the ANCHOR constraints and therefore seems to be a member of a well established family of constraints (McCarthy and Prince, 1995). However, ANCHOR constraints express requirements on the edges of constituents. Ussishkin’s constraint crucially cannot refer to the right edge of the input but has to refer to rightmost consonant of the input. His constraint is therefore not really an ANCHOR constraint. In fact, his constraint can be seen as a constraint requiring that the stem ends in a consonant, and stipulating where this consonant comes from. It is therefore a combination of FINAL-C (requiring a consonant at the edge of the stem), and the ranking ||DEP o UNIFORMITY|| (stating that spreading is preferred over epenthesis).

Hebrew is characterized by the vocalic melody *i-e* in the perfect form (this conjugation is known as the *pi‘el* in the Hebrew grammatical tradition). The vowels /i-e/ are therefore part of the input of a *pi‘el* verb in the perfect. (These vowels can be seen as an inflectional morpheme marking a specific form of the verb.) The input will also contain the consonantal verbal root.[15] The input to form a *pi‘el* perfect of the made-up bi-consonantal root /Q-S/ will then be /Q-S, i-e/, and this input has to map onto the output [QiSeS]. In the observed output form the two [S]’s stand in correspondence to a single input /S/. Such a one-to-many correspondence violates the anti-spreading faithfulness constraint INTEGRITY (McCarthy and Prince, 1995). In order to avoid violation of FINAL-C Hebrew opts to violate INTEGRITY. This gives us evidence for the ranking ||FINAL-C o INTEGRITY||. This ranking is motivated in the tableau in (3).

(3) **||FINAL-C o INTEGRITY||**

/Q-S, i-e/	FINAL-C	INTEGRITY
L Qi.SeS		*
Qi.Se	*!	

This shows why the faithful candidate is not the optimal candidate. However, there are more candidates than just [QiSeS] that avoid violation of FINAL-C, and we

---

[15] Both Bat-El (1994) and Ussishkin (1999) argue against the existence of the bare consonantal root. However, there are reasons to believe that bare consonantal roots do exist. (i) Berent *et al.* (2001a) required of the subjects in their experiment to conjugate bare consonantal roots. Their subjects had no difficulty in doing this. (ii) Berent *et al.* (2001a, 2001b, 2002, Berent and Shimron, 1997) and Frisch and Zawaydeh (2001) found evidence for the fact that restrictions that apply specifically to the consonants of the root have psychological reality. This shows that speakers of Hebrew and Arabic do represent the consonantal root as a separate entity.

It is possible that in denominal verbal formation (the focus of Bat-El’s and Ussishkin’s discussions), the vowels of the nominal root do also form part of the input to the verbal formation process. However, when Hebrew speakers form verbs from roots that are not derived from existing words, and especially when they form verbs from made-up roots, they most likely represent the root as a sequence of bare consonants.

should also rule out these candidates. In the next few paragraphs I will discuss the most important unfaithful competitors of the desired output [QiSeS].

First, we need to account for the fact that FINAL-C is not satisfied by epenthesis of a consonant, i.e. /Q-S, i-e/ does not map onto [QiSeT] with and epenthetic [T]. This candidate will violate the constraint against consonantal epenthesis DEP-C. We also need to consider candidates with different arrangements of the consonants and vowels, i.e. [i.QeS] or [Qi.eS]. These candidates all have onsetless syllables and therefore incur violations of ONSET. Both of these alternative candidates can be ruled out by ranking DEP-C and ONSET higher than INTEGRITY which is violated by the actual output [QiSeS]. This is shown in the tableau in (4).[16]

(4) **||ONSET o DEP-C o INTEGRITY||**[17]

/Q-S, i-e/	ONSET	DEP-C	INTEGRITY
L Qi.SeS			*
QiSeT		*!	
Qi.eS	*!		

Having established the reason for the spreading of one of the root consonants, we now need to establish why it is the final rather than the initial root consonant that spreads – i.e. why does /Q-S, i-e/ not map onto [QiQeS]? I will call on a member of the ALIGNMENT family of constraints (McCarthy and Prince, 1993b) to explain this fact. In

---

[16] Violation of FINAL-C can also be avoided by deletion of one of the input vowels, i.e. /Q-S, i-e/ → [QiS] or [QeS]. This candidate violates MAX-V or some special version of this constraint against deletion from an affix (Gafos, 1998, 2003, McCarthy and Prince, 1995:370, Ussishkin, 1999:417). These candidates can therefore be eliminated by the constraint MAX-V. However, in the discussion in the text I am focusing on unfaithfulness to the root rather than the affixes, and I will therefore not consider these candidates.

[17] We have no evidence for the ranking between ONSET and DEP-C. However, I follow the principle of ranking conservatism (Chapter 3 §2.3 footnote 9, Chapter 4 §2.1.1), and therefore I rank the markedness constraint ONSET over the faithfulness constraint DEP-C.

particular, on a constraint that requires the left edge of the root to coincide with the left edge of the stem. This constraint is defined in (5).

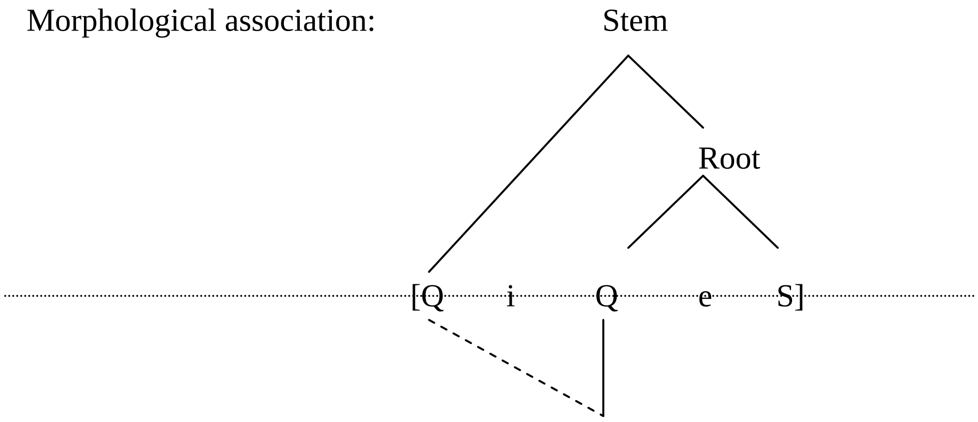
(5) **ALIGN(Root, L, Stem, L)**

The left edge of every root must coincide with the left edge of some stem.

I assume that morphological association and phonological association do not necessarily coincide. Even though the initial [Q] in the output form [QiQeS] is clearly phonologically associated with the /Q/ from the input, it is not morphologically associated with the root morpheme. The initial [Q] in [QiQeS] is a direct daughter of the stem, while the second [Q] is a daughter of the root. The morphological and phonological structure of [QiQeS] is shown in (6).

(6) **Morphological and phonological association in the ungrammatical [QiQeS]**

Morphological association:



Phonological association:

The form [QiQeS] in (6) violates ALIGN(Root, L, Stem, L) (ALIGNL for short) which explains why it is not the observed output. Since both the observed [QiSeS] and the non-observed [QiQeS] violate INTEGRITY, we cannot establish a ranking between ALIGNL and INTEGRITY. However, following the principle of ranking conservatism

(Chapter 4 §2.1.1) I am ranking the markedness constraint ALIGNL higher than the faithfulness constraint INTEGRITY. This is shown in the tableau in (7). In this and all further tableaux I mark morphological association to the root with a superscripted "R" and phonological relatedness with subscripted indexes. Stem boundaries are marked by vertical lines |.

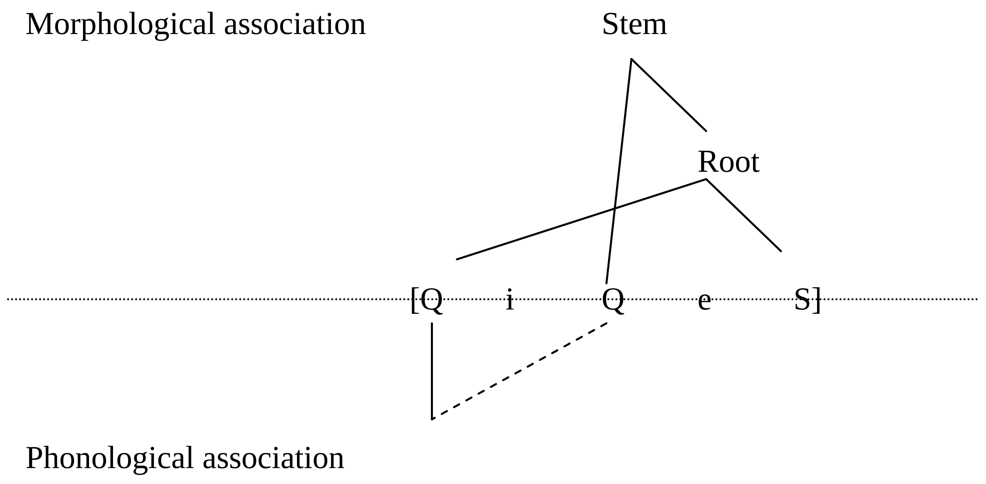
(7) **||ALIGNL o INTEGRITY||**

/Q-S, i-e/	ALIGNL	INTEGRITY
L $\|Q^R\ i\ S_i^R\ e\ S_i\|$		*
$\|Q_i\ i\ Q_i^R\ e\ S^R\|$	*!	*

There are two candidates that obey both FINALC and ALIGNL and that also need to be eliminated. The first root consonant can associate to the left edge of the stem and then to spread onto the second consonantal position of the stem. Similarly, the final root consonant can associate with the right edge of the stem and spread onto the second stem position. These two candidates are represented graphically in (8).

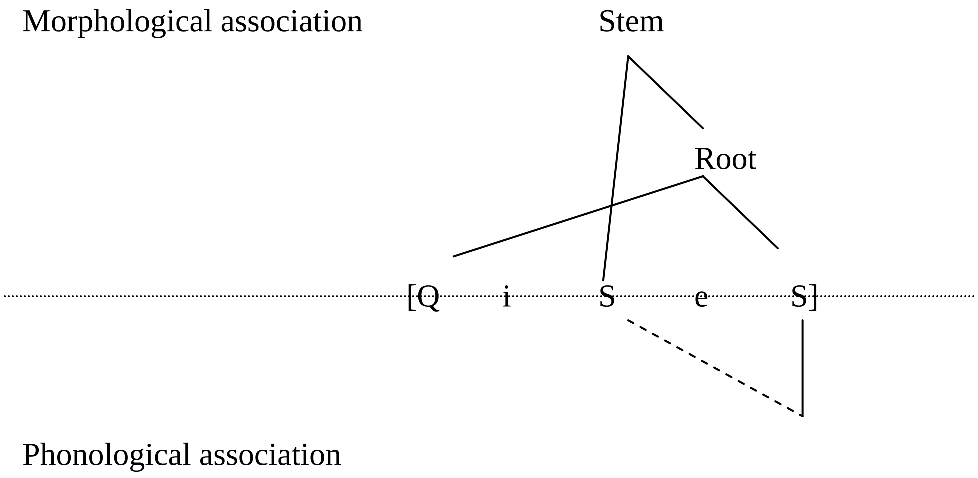
(8) **Two unattested candidates**

a. **Initial root consonant spreads**



((8) continued)

b. **Final root consonant spreads**



Both of these candidates and the observed output candidate [$|Q^{R}$ i S_i^{R} e $S_i|$] violate Integrity. However, both of the forms in (8) also violate the constraint Contiguity (McCarthy and Prince, 1995) indexed to the root – in both of these forms the output stretch that stands in correspondence to the root is discontinuous. This additional violation of Contiguity$_{\text{Root}}$ is what is responsible for eliminating these candidates.[18] Since both the observed and the non-observed candidates violate Integrity, we cannot establish a ranking between Integrity and Contiguity$_{\text{Root}}$. This is shown in the tableau in (9).

(9) **The need for Contiguity$_{\text{Root}}$**

/Q-S, i-e/	Contiguity$_{\text{Root}}$	Integrity
☞ $\|Q^{R}$ i S_i^{R} e $S_i\|$		*
$\|Q^{R}_i$ i Q_i e $S^{R}\|$	*!	*

---

[18] These forms cannot be ruled out by a constraint against crossing association lines. Roots and stems are presumably on separate tiers, so that these structures can be formed without having the root and stem association lines cross.

In the table in (10) I summarize the rankings that I have argued for in this section. The table includes the ranking and a motivation for the ranking. The number in the last column refers to the crucial examples or the ranking argument that motivates the ranking.

(10) **Summary of the rankings thus far**

Ranking	**Motivation**	**Where?**
FINAL-C o INTEGRITY	Final consonant in bi-consonantal roots doubles	(3)
DEP-C o INTEGRITY	FINAL-C satisfied by spreading, not epenthesis	(4)
ONSET o INTEGRITY	FINAL-C satisfied by spreading, not re-arranging vowels and consonants	(4)
FINAL-C, ONSET, ALIGNL o DEP-C, INTEG, CONTIG$_{\text{ROOT}}$	Ranking conservatism	

2.1.3 The OCP: banning identical consonants

In this section I will first show that there is a constraint that bans identical contiguous consonants from the surface realization of a root (§2.1.3.1). After that I will argue that there is also a constraint against contiguous identical consonants in the stem, but that Hebrew tolerates violation of this constraint (§2.1.3.2).

2.1.3.1 No identical consonants in the surface realization of the root

The tendency to ban co-occurrence of identical structures within a specific domain is well established cross-linguistically. Leben (1973) and Goldsmith (1976) originally formulated the Obligatory Contour Principle (OCP) to account for the avoidance of identical contiguous tones. However, the application of the OCP has since been extended also to segmental phonology (McCarthy, 1979, 1986, Padgett, 1991, Yip, 1988).

I will argue here that a specific instantiation of the OCP is responsible for banning contiguous identical consonants from roots in Hebrew. This constraint is defined in (11).

(11) **OCP$_{\mathbf{Root}}$**

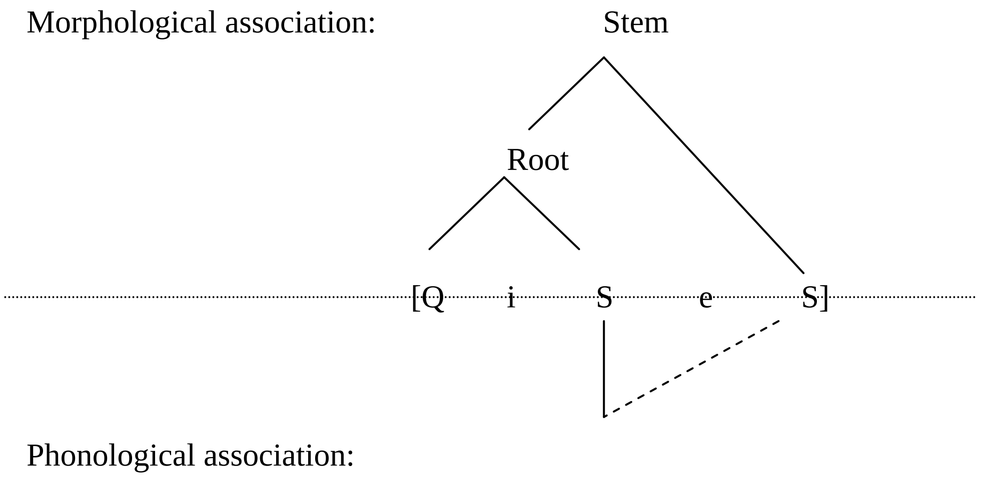
Do not allow identical contiguous consonants in the surface realization of a root.

Note that this constraint is formulated such that it is evaluated on the surface – it does not evaluate the underlying representation of a morpheme, but rather the surface realization of a morpheme. It is like all other markedness constraints in OT a surface oriented constraint. This is an important distinction. Had the constraint been formulated to ban identical contiguous consonants from the underlying representation of a morpheme, it would have been in effect a morpheme structure constraint (Chomsky and Halle, 1968), placing a limitation on what can count as a possible input to the grammar. This would not have been in agreement with the principle of "richness of the base" (Prince and Smolensky, 1993, Smolensky, 1996), one of the central tenets of OT. Richness of the base requires that the grammar be able to handle any input. As I will illustrate below (§2.1.6), OCP$_{\text{Root}}$ will via lexicon optimization (Prince and Smolensky, 1993) result in a lexicon that contains no roots with identical contiguous consonants. But the grammar will still be able to map a hypothetical morpheme with identical contiguous consonants onto a licit output.

It is important that the domain over which the contiguity of two consonants is determined should be the root. Root consonants are often separated from each other by intervening non-root vowels. If contiguity were determined directly over the surface form, then two root consonants separated by non-root vowels would not be contiguous. In a form such as [$Q^{R}iS^{R}eS^{R}$], the two [S]'s are contiguous in the root even if they are not directly contiguous on the surface. This form thus violates OCP$_{\text{Root}}$ even though the two [S]'s are not strictly contiguous.

What is the evidence that the domain over which the Hebrew constraint is evaluated is the root rather than some other morphological or prosodic domain? One part of the evidence rests on the assumption that the phonological and morphological structures of geminate verbs are not isomorphic. Identical consonants do occur in a geminate verb, but the second of the identical consonants is morphologically associated with the stem and not with the root. A form such as [QiSeS] then has the morpho-phonological structure shown in (12). This form does not contain identical consonants within the surface realization of the root. It does contain identical consonants within the surface realization of the stem. Identical consonants are therefore tolerated in the stem.

(12) **Morphological and phonological association of the final-geminate [QiSeS]**



In addition to allowing identical contiguous consonants within the stem, Hebrew also allows identical contiguous consonants within the word. I assume that the stem is the structure to which prefixes and suffixes attach. In (13) I give some examples of identical consonants that are separated by a stem boundary (stem boundaries are indicated by a vertical line |).

(13) **Identical consonants tolerated across a stem boundary in Hebrew**

/ti	+	tfor/	→	[ṯi|ṯfor|]	‘she will sew’	
3 f.s.		sew				(Berent *et al*., 2001a:24)
/li	+	lboš/	→	[li|lboš|]	‘to wear’	
inf. marker		wear				(Berent *et al*., 2001a:24)
/rokem	+	im/	→	[|rokm|im]	‘they sew’	
sew		m. pl.				(Uri Strauss, p.c.)

Contiguous identical consonants are disallowed in the root, but tolerated in the stem and in the word. This serves as motivation for the fact that the domain of the relevant OCP-constraint is indeed the root.

Where does the constraint OCP_{Root} fit into the constraint hierarchy for Hebrew? Suppose that there were a root with two initial identical consonants in its underlying representation, i.e. /Q-Q-S/. Since Hebrew does not tolerate violation of OCP_{Root}, this root cannot be mapped onto a surface realization in which both of the underlying /Q/’s are faithfully preserved. One of the underlying /Q/’s has to be treated unfaithfully. There are several ways in which this can be done. One way is by deleting one of the offending

/Q/’s and then treating the root like a bi-consonantal root – i.e. spreading the final /S/. If this repair is used, then both MAX and INTEGRITY are violated. Both of these faithfulness constraints therefore have to rank lower than OCP$_{\text{Root}}$, i.e. ||OCPRoot o {MAX-C, INTEGRITY}||.[19] This is illustrated in the tableau in (14). I do not consider the vowels in the candidates in this tableau.

(14) **||OCP$_{\text{Root}}$ o {MAX-C, INTEGRITY}||**

/Q-Q-S/	OCP$_{\text{Root}}$	MAX-C	INTEGRITY
$Q^R Q^R S^R$	*!		
L $Q^R S_i^R S_i$		*	*

The tableau in (14) shows that OCP$_{\text{Root}}$ is not a morpheme structure condition in the traditional sense of the word (Chomsky and Halle, 1968). It does not limit what counts as a possible input to the grammar. A root with identical contiguous consonants in underlying form is allowed. However, such a morpheme will never be mapped faithfully onto the surface.

2.1.3.2 Avoiding identical consonants in the stem

If the OCP can be indexed to one morphological category (the root), then it should in principle be possible to index it to other morphological categories such as the stem and the word. In addition to the constraint OCP$_{\text{Root}}$ defined above, there will then also be

---

[19] This is, of course, not the only way in which violation of OCP$_{\text{Root}}$ can be avoided. It is also possible that one of the offending /Q/’s is simply parsed unfaithfully by changing its place of articulation, i.e. /Q/ → [T]. This will then result in a mapping /Q-Q-S/ → [T-Q-S] that also does not violate OCP$_{\text{Root}}$. This would require the ranking ||OCP$_{\text{Root}}$ o IDENT(place)||.

Since the Hebrew lexicon does not contain a root like /Q-Q-S/, we cannot know for sure which of the possible repairs Hebrew would use. In the discussion here I am assuming the deletion-plus-spreading repair, simply because I already need spreading as a way in which to satisfy FINAL-C. However, this is to some extent an arbitrary choice. The same point could be made by assuming the repair used to avoid a violation of OCP$_{\text{Stem}}$ is violation of an IDENT constraint.

other OCP-constraints such as OCP$_{\text{Stem}}$ and OCP$_{\text{Word}}$.[20] I will use OCP$_{\text{Stem}}$ as an example in the discussion below.

(15) **OCP$_{\text{Stem}}$**
Do not allow identical contiguous consonants in the surface realization of a stem.

The ranking of OCP$_{\text{Root}}$ in Hebrew has been determined above. But what about OCP$_{\text{Stem}}$? Unlike OCP$_{\text{Root}}$, Hebrew does tolerate violation of OCP$_{\text{Stem}}$. This rests on the assumption that the morphological and phonological structure of final-geminate verbs are as represented in (12) above. The identical consonants in final-geminates are both affiliated with the stem, and these final-geminates therefore violate OCP$_{\text{Stem}}$.

Since Hebrew tolerates violation of OCP$_{\text{Stem}}$ but not of OCP$_{\text{Root}}$, we can infer the ranking ||OCP$_{\text{Root}}$ o OCP$_{\text{Stem}}$||. From the fact that Hebrew tolerates violation of OCP$_{\text{Stem}}$, we can also infer that all faithfulness constraints that could be violated to avoid a violation of OCP$_{\text{Stem}}$ should rank higher than OCP$_{\text{Stem}}$. Of the faithfulness constraints considered above only DEP-C can be violated in order to avoid an OCP$_{\text{Stem}}$ violation. Suppose that FINAL-C is satisfied not by spreading but by epenthesis of a consonant. The input /Q-S, i-e/ will then map onto a candidate [QiSeT] where the [T] is not phonologically related to any of the consonants in the root. The structure of this candidate is shown in (16).

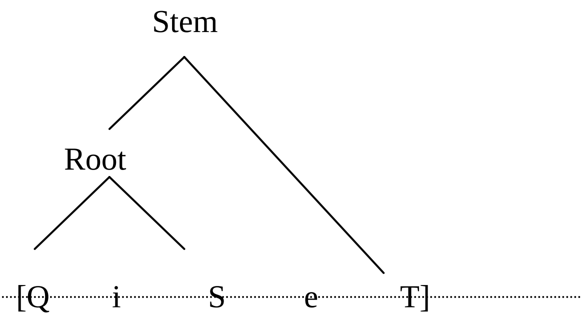
Since Hebrew tolerates violation of OCP$_{\text{Stem}}$, this implies that DEP-C should outrank OCP$_{\text{Stem}}$, i.e. ||DEP-C o OCP$_{\text{Stem}}$||.[21] This is shown in the tableau in (17).

---

[20] There is considerable cross-linguistic evidence for consonantal co-occurrence constraints that apply to different morphological domains. See Tessier (2003, 2004) for a recent review of the data.

(16) **Epenthesis to avoid violation of OCP$_{\text{Stem}}$**

Morphological association:



Phonological association:

(17) **||DEP-C o OCP$_{\text{Stem}}$||**

/Q-S/	DEP-C	OCP$_{\text{Stem}}$
$\|Q^R S^R S\|$		*
$\|Q^R S^R T\|$	*!	▨

The existence of OCP$_{\text{Stem}}$ means that final-geminate verbs, even though they are tolerated in Hebrew, are marked relative to verbs derived from roots with three different consonants. A mapping such as /Q-S-M, i-e/ → [|QiSeM|] does not violate OCP$_{\text{Stem}}$, while a mapping such as /Q-S, i-e/ → [|QiSeS|] does. We should see the effects of OCP$_{\text{Stem}}$ *inter alia* in the processing of non-words. Even tough both [|QiSeM|] and [|QiSeS|] are possible words of Hebrew, [|QiSeS|] is more marked, and should therefore be rated as less well-formed and should be rejected more quickly as a non-word in a lexical decision task.

---

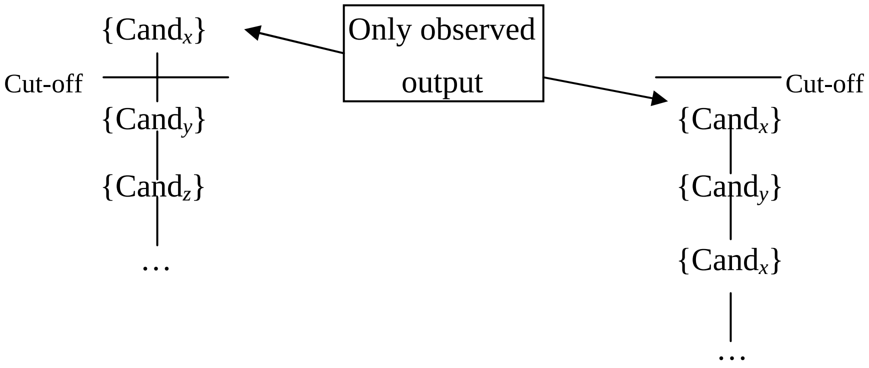
[21] Strictly speaking there is also a situation in which MAX-C in conjunction with DEP-C could be violated in order to avoid a violation of OCP$_{\text{Stem}}$. Consider a root /Q-S-S/ that is mapped onto [$Q^R S^R T$]. In this output, one of the offending /S/'s has deleted in violation of MAX-C. A [T] has then been inserted in violation of DEP-C. However, this is already ruled out but the ranking ||DEP-C o OCP$_{\text{Stem}}$||.

2.1.4 The critical cut-off

The basic theoretical claim that I make in this dissertation is that EVAL does more than just to distinguish between the winning candidate and the mass of losers. EVAL imposes a harmonic rank-ordering on the full candidate set. Language users can then access the full candidate set via this rank-ordering. This idea was used to explain variation in phonological production (Chapters 3, 4 and 5). The most frequently observed variant corresponds to the candidate occupying the highest slot in the rank-ordering, the second most frequent variant to the candidate in the second slot of the rank-ordering, etc. However, if language users have access to the full rank-ordered candidate set, why is variation limited? For most inputs no variation is observed. And even in those instances where variation is observed, it is usually strictly limited to two or three variants. In order to account for the strict limits on variation, the concept of the critical cut-off was introduced (Chapter 1 §2.2.3). The critical cut-off is a point on the constraint hierarchy. If given a choice, language users will not access candidates eliminated by constraints ranked higher than the critical cut-off. Variation is limited because most candidates are eliminated by constraints ranked above the cut-off.

But what of a system with no variation? There are two ways in which an input can be invariantly mapped onto a single output. Suppose that for some input all output candidates except for one is eliminated by a constraint ranked higher than the cut-off. The single candidate that is not eliminated before the cut-off is reached will then be the only observed output candidate – this is just variation with a single observed variant. But there is another way in which an input can be mapped onto a single output. If all candidates violate at least one constraint ranked higher than the cut-off, then language users will be forced to access candidates eliminated by constraints ranked above the cut-off. However, in order to minimize access of candidates eliminated above this cut-off, language users will access only one candidate, namely the candidate occupying the top slot in the rank-ordering. The two ways in which a categorical phenomenon can be modeled are represented graphically in (18). In this representation the rank-ordered candidate set is represented with the candidate rated best at the top. A horizontal line indicates the location of the critical cut-off. Candidates below this line are eliminated by constraints ranked higher than the cut-off.

(18) **Categorical phenomena in a rank-ordering model of EVAL**



This example shows that one way in which to achieve a situation with no variation, is to rank all constraints above the critical-cut-off. All candidates are then guaranteed to violate at least some constraint ranked higher than the cut-off. (This scenario is represented on the right in (18).) Since variation in production is the exception rather the rule, the assumption should be that our grammar should be constructed so as to allow for variation only when there is explicit evidence of variation. For this reason, I made the conservative assumption that the critical cut-off is located as low as possible on the constraint hierarchy when I discussed variation in production (Chapter 4 §3.1).

We have no evidence of variation in the production of Hebrew geminate verbs. Based on the discussion in the previous paragraph it therefore follows that the critical cut-off should be located right at the bottom of the hierarchy for Hebrew. For any input, all output candidates will then violate some constraint ranked higher than the cut-off, so that only one candidate will be accessed for any input.

2.1.5 Summary

In the table in (19) I list all the rankings that I have argued for thus far. Part of this table is a repetition of the table in (9). However, it also contains information in the rankings that were introduced since (9). The table contains the ranking, a short motivation for the ranking, and where in the preceding discussion that ranking was discussed. After the table I give a graphic representation of the rankings.

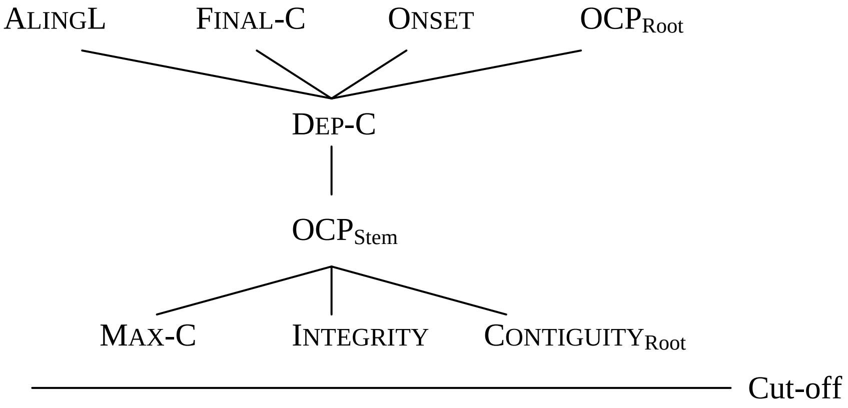
(19) **Summary thus far**

Ranking	**Motivation**	**Where?**
FINAL-C o INTEGRITY	Final consonant in bi-consonantal roots doubles	(3)
DEP-C o INTEGRITY	FINAL-C satisfied by spreading, not epenthesis	(4)
ONSET o INTEGRITY	FINAL-C satisfied by spreading, not re-arranging vowels and consonants	(4)
FINAL-C, ONSET, ALIGNL o DEP-C, INTEGRITY	Ranking conservatism	
OCP$_{Root}$ o MAX-C, INTEGRITY	Geminate roots unfaithfully mapped	(14)
OCP$_{Root}$ o OCP$_{Stem}$	OCP$_{Stem}$ violated but OCP$_{Root}$ not	§2.1.3.2

((19) continued)

Ranking	**Motivation**	**Where?**
DEP-C o OCP$_{Stem}$	OCP$_{Stem}$ violation not avoided by epenthesis	(17)
OCP$_{Root}$ o DEP-C	Ranking conservatism	
OCP$_{Stem}$ o INTEG, MAX-C, CONTIG$_{Root}$	Ranking conservatism	
All constraints o Cut-off	No variation, ranking conservatism	§2.1.4

(20) **Graphic representation of the rankings for Hebrew**



In the rest of this section I will illustrate how the hierarchy in (20) generates only actually observed forms for Hebrew verbs. In the next section (§2.1.6), I will then show how this hierarchy together with lexicon optimization ensures that the Hebrew lexicon will not contain roots with identical contiguous consonants.

Let us begin by considering the simplest case – a verbal root with three non-identical consonants. In such a form all three root consonants should be faithfully parsed onto the surface, i.e. /Q-S-M, i-e/ → [QiSeM]. The tableau in (21) considers this case. In this and all further tableaux I do not consider the vowels. The tableaux from here on should also be interpreted like tableaux in the rank-ordering model of EVAL – see Chapter 1 §2.2.1 and §2.2.3 for the conventions used in these tableaux.

(21) **/Q-S-M/ → [Q S M]**

/Q-S-M/	ALIGNL	FINAL-C	ONSET	OCP_{Root}	DEP-C	OCP_{Stem}	INTEGRITY	MAX-C	$\text{CONTIG}_{\text{Root}}$
1 $|Q^R\ S^R\ M^R|$									
3 $|Q^R\ S^R\ T|$	*!				*			*	
2 $|Q^R S^R M^R_i M_i|$	*!					*	*		

Output of EVAL[22]

Cut-off

L $|Q^R\ S^R\ M^R|$

Other candidates Some faithfulness constraint

Since the faithful candidate does not violate any of the markedness constraints, any faithfulness violation is guaranteed to be fatal. All candidates appear below the line representing the critical cut-off. Only one candidate is therefore accessed, namely the faithful candidate.

Now consider how a bi-consonantal root will be handled in this grammar. Such an input has to be mapped onto a form with a final-geminate, i.e. /Q-S, i-e/ → [QiSeS]. The tableau in (22) shows that our grammar does indeed select the candidate as output for such as input.

---

[22] Candidate [$|Q^R\ S^R\ M^R|$] violates none of the constraints in this mini-grammar. However, it does violate several constraints that are just not considered here. This is why this candidate still occurs below the critical cut-off on this representation.

(22) **/Q-S / → [Q S S]**

/Q-S/		ALIGNL	FINAL-C	ONSET	OCP_{Root}	DEP-C	OCP_{Stem}	INTEGRITY	MAX-C	$\text{CONTIG}_{\text{Root}}$
1	$\|Q^R\ S_i^R S_i\|$						*	*		
5	$\|Q_i\ Q_i^R\ S^R\|$	*!					*	*		
4	$\|Q^R\ S^R\|$		(*)[23]	(*)[23]						
3	$\|Q^R\ S^R\ T\|$					*!				
2	$\|Q^R\ S_i\ S_i^R\|$						*	*		*!

Output of EVAL

———————— Cut-off

L $|Q^R\ S_i^R S_i|$

|

$|Q^R\ S_i\ S_i^R|$ $_{\text{CONTIGUITYRoot}}$

|

Other candidates $_{\text{DEP-C or higher}}$

In the second candidate [$|Q_i\ Q_i^R\ S^R|$] the root and stem are misaligned at their left edges. This leads to a fatal violation of ALIGNL. In the third candidate [$|Q^R\ S^R|$] there are not enough consonants to satisfy both FINAL-C and ONSET simultaneously. This candidate actually stands in for two candidates here, namely [$|Q^R$ i . S^R e|] and [Q^R i . e $S^R|$]. The first violates FINAL-C and the second ONSET. The fourth candidate [$|Q^R\ S^R\ T|$] inserts a consonant, thereby earning a violation of DEP-C. The last candidate $|Q^R\ S_i\ S_i^R|$ deletes on the two root /S/'s, and then spreads the remaining /S/ into the middle stem position (this is the candidate from (8b) above). The portion of the surface string that stands in correspondence to the root is not contiguous, which earns this candidate a fatal violation of $\text{CONTIGUITY}_{\text{Root}}$. The first candidate [$|Q^R\ S_i^R S_i|$] is then selected as

---

[23] See the discussion just below the tableau about these violations.

output in spite of its OCP_{Stem} and INTEGRITY violations. All candidates violate at least one constraint ranked higher than the cut-off. Only the single best candidate is therefore predicted to be accessed – i.e. no variation is predicted.

What remains to show now is how this grammar will deal with a root that has two contiguous identical consonants. I will show just below that lexicon optimization will prevent such roots from being added to the lexicon of Hebrew. However, richness of the base requires that the grammar at least be able to handle such inputs. The tableau in (23) shows how the grammar will handle a root with identical second and third consonants.

(23) **/Q-S-S/ → $[\text{Q}^{\text{R}}\ \text{S}_i^{\text{R}}\ \text{S}_i]$**

/Q-S-S/		ALIGNL	FINAL-C	ONSET	OCP_{Root}	DEP-C	OCP_{Stem}	INTEGRITY	MAX-C	$\text{CONTIG}_{\text{Root}}$
1	$\|\text{Q}^{\text{R}}\ \text{S}_i^{\text{R}}\ \text{S}_i\|$						*	*	*	
3	$\|\text{Q}^{\text{R}}\ \text{S}^{\text{R}}\ \text{S}^{\text{R}}\|$				*!		*			
4	$\|\text{Q}_i\ \text{Q}^{\text{R}}\ \text{S}^{\text{R}}\|$	*!					*	*	*	
2	$\|\text{Q}^{\text{R}}\ \text{S}^{\text{R}}\ \text{T}\|$					*!			*	

Output of EVAL

Cut-off

L $\|\text{Q}^{\text{R}}\ \text{S}_i^{\text{R}}\ \text{S}_i\|$ OCPStem

$\|\text{Q}^{\text{R}}\ \text{S}^{\text{R}}\ \text{T}\|$ DEP-C

Other candidates ALIGNL or OCPRoot

The second candidate $[\|\text{Q}^{\text{R}}\ \text{S}^{\text{R}}\ \text{S}^{\text{R}}\|]$ is the faithful candidate. It fatally violates OCP_{Root}. The third candidate $[\|\text{Q}_i\ \text{Q}^{\text{R}}\ \text{S}^{\text{R}}\|]$ avoids violation of OCP_{Root} by deleting one of the offending /S/'s and then spreading the /Q/ into the initial position of the stem. Since this initial [Q] is not morphologically associated with the root, this candidate does not violate OCP$_{\text{Root}}$. However, at the same time the disassociation between the initial [Q] and the root means that the left edges of the root and the stem do not coincide. This candidate therefore fatally violates ALIGNL. The last candidate $[|Q^{R}\ S^{R}\ T|]$ also avoids violation of OCP$_{\text{Root}}$ by deletion one of the two [S]'s. In order to satisfy FINAL-C it then inserts a third consonant that is unrelated to any of the input consonants. This earns it a fatal violation of DEP-C. This leaves only the first candidate $[|Q^{R}\ S_{i}^{R}\ S_{i}|]$. This candidate also avoids violation of OCP$_{\text{Root}}$ by deleting one of the offending /S/'s. However, this candidate spreads the remaining /S/ into the final stem position, earning it a violation of INTEGRITY and OCP$_{\text{Stem}}$. Both of these constraints rank lower the constraints violated by other candidates. The first candidate is therefore rated best by EVAL. Again, since all candidates violate at least one constraint ranked higher than the cut-off, no variation is predicted. The result is that a root with identical contiguous consonants will never be mapped faithfully. The situation is the same with a root with identical first and second consonants.

Our grammar does therefore correctly account for the surface patterns observed in Hebrew. Recall that the goal at the outset of this section was to explain the asymmetry in the distribution of identical consonants in Hebrew verbs. In particular, that final-geminates [QiSeS] are well-formed while initial-geminates [QiQeS] are not. How does the grammar developed here account for this fact? There are two possible sources of geminates – they can either be derived from bi-consonantal roots or from tri-consonantal roots with identical consonants. Consider first the bi-consonantal roots. We have to explain why /Q-S/ maps onto $[|Q^{R}\ S_{i}^{R}\ S_{i}|]$ and not onto $[|Q_{i}\ Q_{i}^{R}\ S^{R}|]$. Both of these forms violate INTEGRITY and OCP$_{\text{Stem}}$. However, the stem and root of $[|Q_{i}\ Q_{i}^{R}\ S^{R}|]$ are misalignment at their left edges, earning this form an additional violation in terms of ALIGNL.

Now consider the second possible source of geminates – tri consonantal roots with contiguous identical consonants (/Q-Q-S/ and /Q-S-S/). Because of the high ranking of OCP$_{Root}$ the surface realization of a root is not allowed to contain two contiguous identical consonants. In order avoid violation of OCP$_{Root}$, one of the identical consonants is deleted – hence the ranking ||OCP$_{Root}$ o MAX-C||. The remaining two root consonants are then treated just like a bi-consonantal root – the final root consonants spreads.

2.1.6 Lexicon optimization

When a Hebrew speaker is presented with a word that has two contiguous identical consonants, what prevents him/her from inferring that this word is derived from a root with identical consonants? More concretely, if a Hebrew speaker were to hear a word like [QiSeS], why will he/she assume that this word is the result of an unfaithful mapping from /Q-S, i-e/ rather than the result of the faithful mapping from /Q-S-S, i-e/? The principle of “lexicon optimization” (Prince and Smolensky, 1993) is responsible for this.

Lexicon optimization is proposal for how language users acquire the entries of their lexicon. The idea is that upon perceiving a word [output], language users consider all possible /input/-forms that could be mapped onto [output] by their grammar. They then select from among all the possible /input/-forms that one that results in the most harmonic /input/ → [output] mapping. In lexicon optimization the output candidate is kept constant and the input is varied. Lexicon optimization is then a comparison between different inputs for the same output candidate rather than a comparison between different output candidates for the same input.

The Hebrew example that we are dealing with here is slightly more complicated than this. The output form [QiSeS] can have more than one morphological and phonological structure. For illustration I list just a few of the more obvious structures: [|Q^R i S^R e S^R|] (perfect alignment or root and stem, completely faithful), [|Q^R i S_i^R e S_i|] (root and stem misaligned at right edge, second [S] is the result of spreading from the first), [|Q^R i S^R e S|] (root and stem misaligned at the right edge, second [S] is result of epenthesis), etc. All of these different morpho-phonological forms have the same phonetic interpretation (are pronounced exactly the same). Upon hearing [QiSeS] a Hebrew speaker therefore has to do more than simply consider different possible inputs for this percept. He/she also has to consider different possible morphological and phonological structures for the percept.

I propose that Hebrew speakers will proceed as follows: First, they will feed /Q-S-S, i-e/ and /Q-S, i-e/ separately through their grammars – i.e. a straightforward production oriented process. From each of these two comparisons they will then select the best /input/ → [output] mapping. Lexicon optimization is achieved by comparing the best /input/ → [output] mappings from each of these two comparisons. In (22) and (23) above I showed how the Hebrew grammar that I have developed will treat a /Q-S, i-e/ and /Q-S-S, i-e/ input. The most harmonic mapping from tableau (22) is /Q-S, i-e/ → [|Q^R i S_i^R e S_i|] and the most harmonic mapping from tableau (23) is /Q-S-S, i-e/ → [|Q^R i S_i^R e S_i|]. Lexicon optimization is performed by comparing these two mappings with each other. This comparison is shown in (24).

(24) **Lexicon optimization: [QiSeS] from /Q-S/ not /Q-S-S/**

		AlignL	Final-C	Onset	OCP$_{Root}$	Dep-C	OCP$_{Stem}$	Integrity	Max-C	Contig$_{Root}$
2	/Q-S-S/ → [|$Q^R\ S_i^R S_i$|]						*	*	*	
1	/Q-S/ → [|$Q^R\ S_i^R S_i$|]						*	*		

Output of EVAL

/Q-S/ → |$Q^R\ S_i^R S_i$|

/Q-S-S/ → |$Q^R\ S_i^R S_i$| Max-C

In the second mapping /Q-S/ → [|$Q^R\ S_i^R S_i$|] the input /S/ spreads, earning this mapping a violation of Integrity. The two surface [S]'s are both affiliated with the stem, so that this mapping also violates OCP$_{Stem}$. In the first mapping /Q-S-S/ → [|$Q^R\ S_i^R S_i$|] the second input /S/ deletes, earning this mapping a violation of Max-C. The remaining /S/ then spreads so that this mapping also violates Integrity. In this mapping the two surface [S]'s are both affiliated with the stem, so that this mapping also violates OCP$_{Stem}$. The extra Max-C violation of the first mapping implies that the second mapping is the more harmonic of the two. This means that upon hearing a word such as [QiSeS], a Hebrew speaker will infer that this word was derived from a bi-consonantal rather than a tri-consonantal root.

The grammar of Hebrew can successfully handle roots with identical contiguous consonants (see the tableau in (23) above). However, due to lexicon optimization Hebrew speakers will never infer the existence of such roots.

2.1.7 The processing of non-words

Now that we have a mini grammar for Hebrew, we can consider how Hebrew speakers will process non-words. In the experiments that I will discuss below (§2.2), Berent *et al*. used three kinds of non-words, namely non-words with three non-identical consonants in the stem positions ([|QiSeM|]), non-words with identical consonants in the last two stem positions ([|QiSeS|]), and non-words with identical consonants in the first two stem positions ([|QiQeS|]).

I will assume that for any geminate non-word percept, listeners have two choices for what the input root for the percept could be. They can assume that the input root is identical to the consonants of the stem in the perceived non-word, or they can assume that the input root is bi-consonantal. For a percept [|QiSeS|] they can therefore assume that the input form of the root is /Q-S-S/ or that it is /Q-S/. For a percept [|QiQeS|] listeners can assume that input form of the root is /Q-Q-S/ or that it is /Q-S/. However, for a non-word percept with three non-identical consonants in stem position, listeners will consider only one possible input for the root, namely a form that is identical to the consonants in the stem of the non-word percept. For a [|QiSeM|] percept, listeners will only consider /Q-S-M/ as input for the root.

(25) **Underlying representation for roots in non-words**

	Percept	**UR considered**
No-geminate:	[QiSeM]	/Q-S-M/
Final-geminate:	[QiSeS]	/Q-S/ and /Q-S-S/
Initial-geminate:	[QiQeS]	/Q-S/ and /Q-Q-S/

The tableau below in (26) compares each of the mappings that listeners will consider when they are processing non-words.

(26) **Comparing non-words**

Geminate?		ALIGNL	FINAL-C	ONSET	OCP_{Root}	DEP-C	OCP_{Stem}	INTEGRITY	MAX-C	$\text{CONTIG}_{\text{Root}}$
No	1 /Q-S-M/ → $[\|Q^RS^RM^R\|]$									
Final	2 /Q-S/ → $[\|Q^RS_i{}^RS_i\|]$						*	*		
	3 /Q-S-S/ → $[\|Q^RS^RS^R\|]$				*		*			
Initial	4 /Q-S/ → $[\|Q_iQ_i{}^RS^R\|]$	*					*	*		
	3 /Q-Q-S/ → $[\|Q^RQ^RS^R\|]$				*		*			

Output of EVAL

Cut-off

/Q-S-M/ → $[|Q^RS^RM^R|]$ No-gemination

|

/Q-S/ → $[|Q^RS_i{}^RS_i|]$ Final-gemination

|

/Q-S-S/ → $[|Q^RS^RS^R|]$ Final/Initial-gemination
/Q-Q-S/ → $[|Q^RQ^RS^R|]$

|

/Q-S/ → $[|Q_iQ_i{}^RS^R|]$ Initial-gemination

This comparison shows the following: (i) Non-words with no-gemination are the most well-formed. (ii) Depending on what underlying representation is assumed, final-gemination non-words are either more well-formed than initial-gemination non-words or equally well-formed with initial-gemination non-words. (iii) Initial-gemination non-words are either less well-formed than final-gemination non-words or equally well-formed with final-gemination non-words.

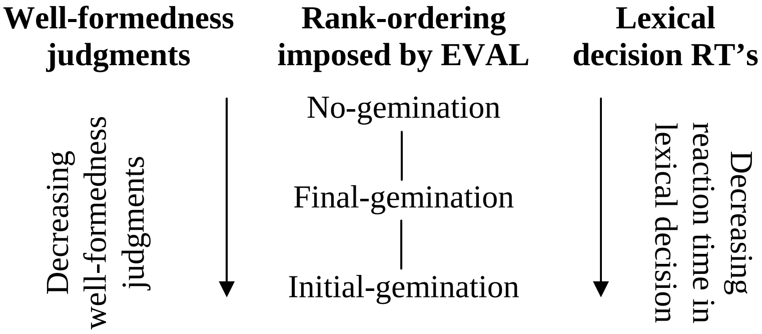
Suppose that every time a listener is presented with a non-word with gemination, he/she considers both potential input forms and selects the form that results in the most harmonic parsing (i.e. performs lexicon optimization every time). The listener will then eventually settle on bi-consonantal input roots for final-geminate forms. The derivation from a bi-consonantal root to final-geminate is more well-formed than the derivation of an initial-geminate form, irrespective of whether the initial-geminate is derived from a bi-consonantal or a tri-consonantal root. The prediction is therefore that final-geminates will be more well-formed than initial-geminates.

(27) **Well-formedness of different kinds of non-words**

No-gemination	[QiSeM]	Most well-formed
Final-gemination	[QiSeS]	
Initial-gemination	[QiQeS]	Least well-formed

If language users use the grammar in the processing of non-words, then we would expect to see effects of this relation in the way in which they process non-words. In particular we would expect that: (i) In well-formedness judgment experiments language users would rate no-gemination non-words better than final-gemination non-words, and final-gemination non-words again better than initial-gemination non-words. (ii) In lexical decision tasks, language users should on average reject initial-gemination non-words more quickly than final-gemination non-words, and final-gemination non-words again more quickly that no-gemination non-words. These predictions are represented graphically in (28). In the next section I discuss a series of experiments performed by Berent *et al*. in which (most of) these predictions are confirmed.

(28) **Predictions with regard to well-formedness judgments and lexical decision**



2.2 Word-likeness and lexical decision in Hebrew

Berent *et al*. (2001a, 2001b, 2002, Berent and Shimron, 1997) performed a series of experiments in which they tested whether the restriction on the distribution of identical consonants in Hebrew roots influences the manner in which speakers of Modern Hebrew process non-words. These experiments included both well-formedness judgment experiments and lexical decision experiments. In the well-formedness judgment experiments, the subjects reacted according to the predictions of the analysis developed above – that is, no-gemination rated best, then final-gemination and then initial-gemination. The results of the lexical decision experiments agree only partially with the predictions. In accordance with the predictions, final-geminate forms were rejected slower than initial-geminate forms. However, contra the predictions, final-geminate forms were also rejected slower than no-geminate forms. (See §2.2.2 below for a discussion of what could have caused the slowdown in reaction times associated with geminate forms.)

In the rest of this section I will first discuss the results of the well-formedness judgment experiments performed by Berent *et al*., and then the results of their lexical decision experiments. I will not discuss their experimental design in detail, for the following two reasons: (i) The information about their experimental design is discussed in detail in the published reports of their research. (ii) Their experiments were designed to test hypotheses that are different from the hypotheses that I discuss here. Many of the details of their experimental design are therefore not relevant to the discussion here.

2.2.1 Well-formedness judgment experiments

Berent *et al*. conducted two kinds of well-formedness judgment experiments, namely gradient well-formedness judgment experiments and comparative well-formedness judgment experiments. In the gradient well-formedness judgment experiments subjects are presented with individual tokens and are required to rate each token on a 5-point scale for its well-formedness. In the comparative well-formedness judgment experiments subjects are presented with three tokens at a time, and they are required to arrange the three tokens in the order of their well-formedness.

2.2.1.1 Gradient well-formedness judgment experiments

There are three separate studies in which Berent *et al*. conducted gradient well-formedness judgment experiments. These are Berent *et al*. (2001a), Berent *et al*. (2002), and Berent and Shimron (1997). There are slight differences in design between the three experiments. However, these differences are small and not relevant to the point of the current discussion. The results in all three of experiments were basically the same. I will therefore discuss one of the three experiments as a representative example. For this purpose I select the results of Berent and Shimron (1997).

Berent and Shimron selected 24 tri-consonantal roots, each consisting of three non-identical consonants. They also selected 24 bi-consonantal roots, each consisting of two non-identical consonants. None of these 48 roots corresponded to actual roots of Hebrew. They conjugated each of the 24 tri-consonantal roots in three different verbal forms, so that there were 72 non-words formed from tri-consonantal roots. The 24 bi-consonantal roots were transformed into initial and a final-geminate stems by doubling either the initial or final consonant of each root. Each of these geminates was then also conjugated in three verbal forms, so that there were 72 initial-geminate non-words and 72 final-geminate non-words. In total, the stimulus list for this experiment consisted of 216 non-words conjugated as verbs.

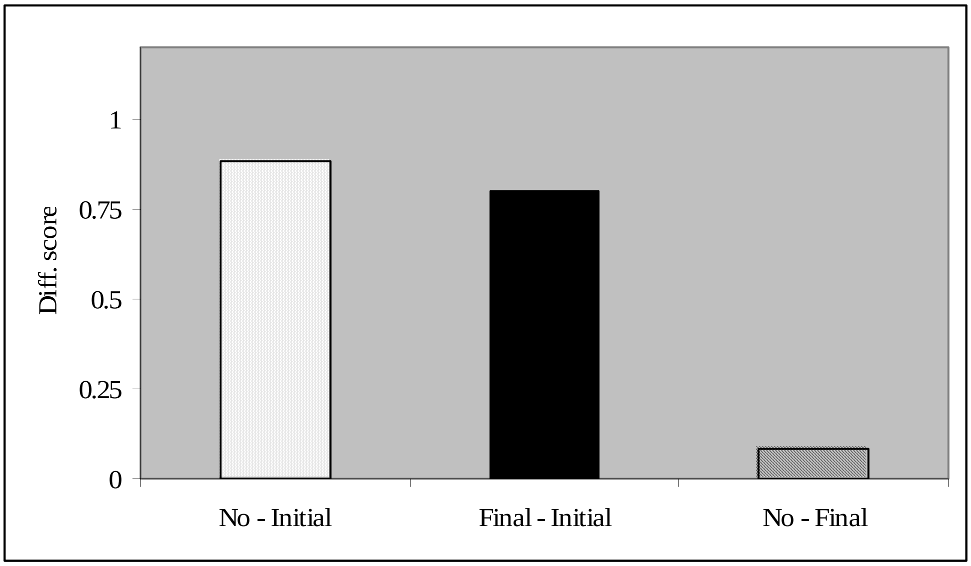
This list of forms was randomized and presented in written form to 15 native speakers of Hebrew. The task of the subjects was to rate each non-word as to its word-likeness or well-formedness. Forms were rated on a 5-point scale, where [5] indicated a form that sounded excellent as a possible word, and [1] a form that sounded impossible as a word of Hebrew. Based on the analysis developed above (§2.1), the expectation is that no-geminate forms would on average be rated better than final-geminate forms, which again would be rated better than initial-geminate forms – the less marked a form, the better it would be rated.

Berent and Shimron do not report the average scores assigned to each of the three token types. However, they do report the difference scores – i.e. the difference between the average ratings assigned to each of the three token types.[24] The results of this experiment are summarized in the table in (29), and represented graphically in (30).

(29) **Average well-formedness ratings in the gradient well-formedness judgment experiment of Berent and Shimron (1997)**

Comparison	Example	Difference score	*t*	*df*	*p*
Initial-gemination and no-gemination	Q-Q-S Q-S-M	0.881	11.139	46	< 0.000
Initial-gemination and final-gemination	Q-Q-S Q-S-S	0.801	9.984	46	< 0.000
Final-gemination and no-gemination	Q-S-S Q-S-M	0.081	_[25]		> 0.05

(30) **Gradient well-formedness ratings: Difference scores between the mean ratings for the different token classes in the Berent and Shimron (1997) data**



---

[24] The difference scores were computed by subtracting the average score for the more marked token type from the average score for the less marked token type – i.e. a positive difference score means that the less marked token type was on average rated better. Specifically: (i) Difference Score (Initial~Final) = Mean Score (Final) – Mean Score (Initial). (ii) Difference Score (Initial~No) = Mean Score (No) – Mean Score (Initial). (iii) Difference Score (Final~No) = Mean Score (No) – Mean Score (Final).

[25] Berent and Shimron unfortunately do not report the *t*-statistic for this comparison. They do, however, report that a *p*-value of larger than 0.05 was obtained for this comparison using the Tukey HSD test – i.e. there was no significant difference between the mean ratings for these two classes of tokens.

In this experiment there is evidence for the fact that initial-geminate forms were rated worse than both final-geminate forms and no-geminate forms. Since forms with initial-gemination are predicted to be the most marked by the analysis developed above, this corresponds to our expectation. But what about the comparison between no-geminate and final-geminate forms? Although both of these classes represent possible words of Hebrew, under the analysis developed above final-geminate forms are more marked than no-geminate forms. The prediction is that final-geminate forms should receive lower scores than no-geminate forms. In absolute terms this was confirmed by the results of the experiment – no-geminate forms were rated on average 0.081 points better than final-geminate forms. However, this is a very small difference and it did not reach the critical level of statistical significance.

What should we make of the failure to find a difference between no-geminate and final-geminate forms? First, it should be noted that the results do not go against the predictions of the analysis developed above in §2.1 – it simply does not confirm these predictions. Secondly, Berent and Shimron claim that the non-result in this condition can be attributed to the experimental design. Both no-geminate and final-geminate forms are possible words of Hebrew. It is then possible that both of these types of words were rated very well, and that there was simply not enough room at the top end of the 5-point scale to differentiate between these two classes of tokens. The non-difference between these two token types can therefore be attributed to a ceiling effect.[26]

---

[26] In order to counter this problem, Berent and Shimron conducted a second type of well-formedness judgment experiment in which subjects had to rate final-geminates and no-geminates in relation to each other – i.e. the experimental design forced a direct comparison between these two kinds of forms. In this comparative well-formedness experiment, they did find evidence for the predicted difference between final-geminate forms and no-geminate forms. See §2.2.1.2 just below.

2.2.1.2 Comparative well-formedness judgment experiments

Berent and Shimron (1997) and Berent *et al.* (2001a, 2002) also conducted comparative well-formedness judgment experiments. Since the results of the three replications of these experiments were basically the same, I discuss only the results of Berent and Shimron (1997) here are a representative example.

In the comparative well-formedness judgment experiment, Berent and Shimron used the same 48 roots (24 tri-consonantal and 24 bi-consonantal) as in the gradient well-formedness judgment experiment. They matched each bi-consonantal root with a tri-consonantal root that differed from the bi-consonantal root by the addition of one consonant. Each pair of bi- and tri-consonantal roots was converted into a triple of tri-consonantal stems by geminating the first and last consonant of the bi-consonantal root. A stem triple consisted of a stem with three non-identical consonants, an initial-geminate and a final-geminate. The tri-consonantal stem shared two of its consonants with the two geminate stems. These triples were conjugated in three different verbal forms of Hebrew, so that each triple corresponded to three triples of non-words conjugated as verbs. A stem triple |Q-S-M|~|Q-S-S|~|Q-Q-S| could, for in stance, correspond to the following three non-word triples: (i) [|QiSeM|]~[|QiSeS|]~[|QiQeS|], (ii) [|maQSiMim|]~[|maQSiSim|]~[|maQQiSim|], (iii) [|hitQaSeMtem|]~[|hitQaSeStem|]~[|hitQaQeStem|]. Each of the 24 triples corresponded to three such non-word triples, for a total of 72 non-word triples.

These 72 non-word triples were randomized and presented to 18 native speakers of Hebrew in written form. Subjects had to order the members of each triple according to their acceptability as words of Hebrew. A score of [3] was assigned to the member that was considered to be most word-like, a score of [1] to the member that was considered least word-like, and a score of [2] to the remaining member. This experimental design results in a direct comparison between no-gemination, final-gemination and initial-gemination forms. Based on the analysis developed above (§2.1), we expect no-geminate forms to be rated better than final-geminate forms, which again would be rated better than initial-geminate forms – the less marked a form, the better it would be rated.

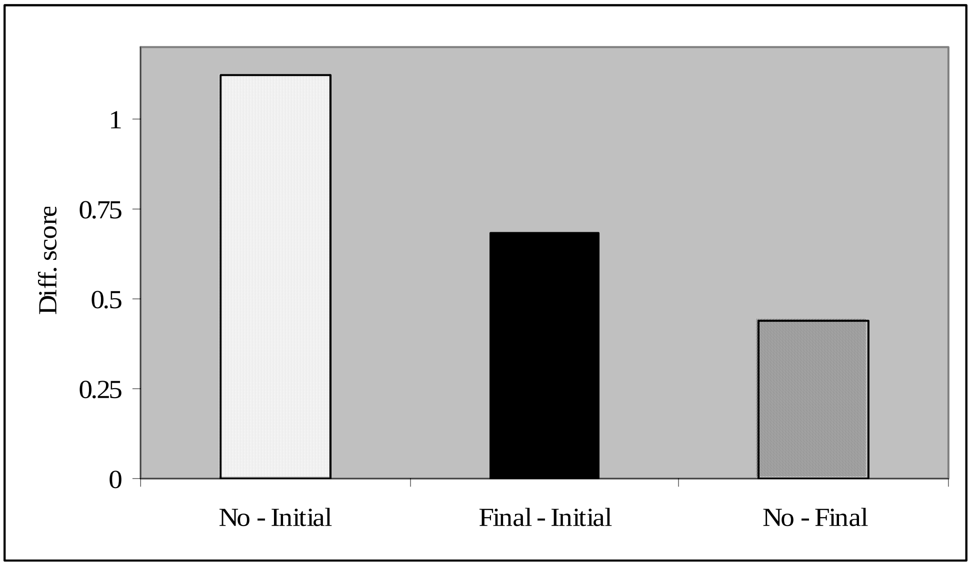
As in the gradient well-formed judgment experiment, Berent and Shimron do not report the average scores assigned to each of the three token types, but only the difference scores between the three token types. The results of this experiment are summarized in the table in (31), and represented graphically in (32).

In accordance with the predictions of the analysis in §2.1, the initial-geminate forms were rated worse than both the final-geminate forms and the no-geminate forms. This corresponds to the fact that initial-geminate forms are predicted to be more marked than both final-geminate forms and no-geminate forms. In the gradient well-formedness judgment experiment no difference was found between the two kinds of possible words, i.e. between final-geminate forms and no-geminate forms. That null result was attributed to the experimental design – no direct comparison was possible between these two kinds of tokens and they received equally good ratings as possible words. In the comparative well-formedness experiment, subjects are required by the experimental design to compare these two forms directly. If there is a difference between these two kinds of tokens, we expect to find evidence of that difference in this experiment. And no-geminate forms were indeed preferred over final-geminate forms in this experiment. Even though both final-geminate and no-geminate forms are possible words of Hebrew, Hebrew speakers consider no-geminate forms as more word-like/well-formed than final-geminate forms.

(31) **Average well-formedness ratings in the comparative well-formedness judgment experiment of Berent and Shimron (1997)**

Comparison	**Example**	**Difference score**	***t***	***df***	***p***
Initial-gemination and no-gemination	Q-Q-S Q-S-M	1.122	18.55	46	< 0.000
Initial-gemination and final-gemination	Q-Q-S Q-S-S	0.682	11.28	46	< 0.000
Final-gemination and no-gemination	Q-S-S Q-S-M	0.44	_[27]		< 0.05

(32) **Comparative well-formedness: Difference scores between the mean ratings for the different token classes in the Berent and Shimron (1997) data**



This experiment confirms the predictions of the analysis developed in §2.1. In terms of well-formedness judgment experiments we have no evidence against the predictions of the analysis, and some evidence in favor of this analysis. Consequently, I interpret the results of the well-formedness judgment experiments as confirming the predictions of the analysis developed in §2.1 above. The less marked a token is according to the grammar, the more well-formed/word-like it is judged to be by language users.

---

[27] Berent and Shimron do not report the *t*-score for this comparison. They do report that a *p*-value of smaller than 0.05 was obtained for this comparison using the Tukey HSD test.

2.2.2 Lexical decision experiments

Both Berent *et al*. (2001b) and Berent *et al*. (2002) performed experiments to determine whether the restriction on the distribution of identical consonants influences the manner in which Hebrew speakers perform lexical decision tasks. The results of all of these experiments were basically the same, and I will therefore discuss one of the experiments as representative. For this purpose I select *Experiment 2* of Berent *et al*. (2001b).

Berent *et al*. selected 30 tri-consonantal roots, each consisting of three non-identical consonants. They also selected 30 bi-consonantal roots, each consisting of non-identical consonants. Not one of these roots corresponded to actual roots of Hebrew. They transformed each of the bi-consonantal roots into two tri-consonantal stems by geminating either the first or the last consonant of the root. This resulted in 90 stems. Of these 90 stems, 30 had no identical consonants (no-gemination), 30 had identical consonants in the initial two positions (initial-gemination), and 30 had identical consonants in the final two positions (final-gemination). Each of these stems was inserted into the same verbal pattern of Hebrew, resulting in 90 non-words, conjugated as verbs.

Berent *et al.* also selected 45 tri-consonantal and 45 bi-consonantal Hebrew roots. These roots all corresponded to actual Hebrew verbal roots. These 90 roots were inserted into the same verbal forms as the non-words, resulting in 90 actual Hebrew roots. The 90 words and 90 non-words were randomized into one list. The list was presented in a lexical decision task to 20 native speakers of Hebrew. The tokens were presented one at a time on a computer monitor. Subjects had to indicate whether the token was or was not a word of Hebrew by pushing one of two buttons. Response times were recorded. Statistical analyses were done on the response times to correct non-word responses.

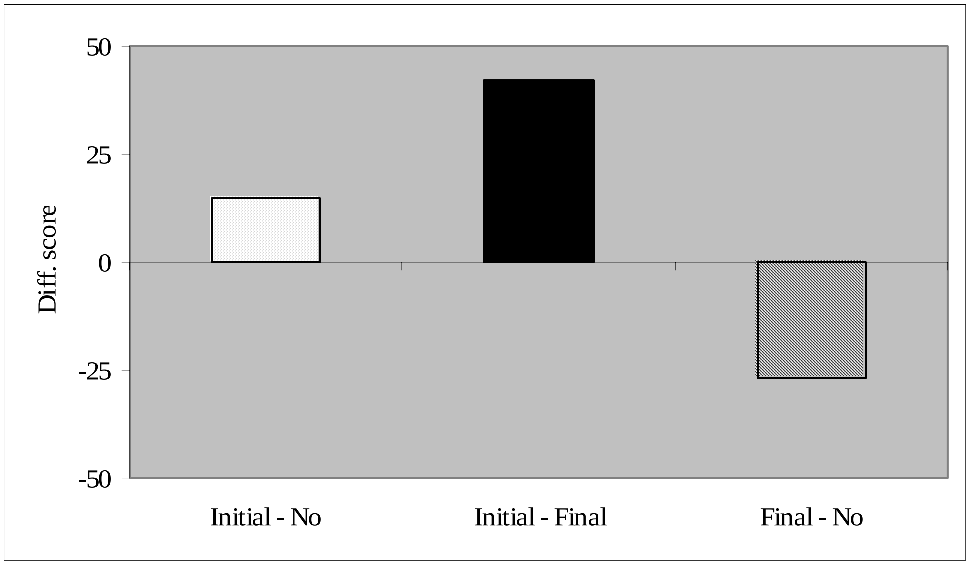
Recall the basic prediction with regard to lexical decision – the more well-formed a non-word token is according to the grammar, the more seriously the language user will consider it as an actual word of Hebrew. Consequently, we expect slower reaction times with more well-formed non-words than with less well-formed non-words. In terms of the analysis developed above in §2.1, we therefore expect that initial-geminate forms should be rejected the most quickly, final-geminate forms more slowly, and no-geminate forms the most slowly. The results of this experiment confirmed some of these predictions. In particular, the results confirmed the prediction that initial-geminate forms should be rejected more quickly than final-geminate forms. However, final-geminate forms were not rejected more quickly than no-geminate forms. In fact, the opposite was found – final-geminate forms were rejected more slowly than no-geminate forms. The table in (33) summarizes the results, and (34) contains a graphic representation of the results.

In the analysis developed in §2.1, final-geminate forms are less well-formed than no-geminate forms. Under the assumption that language users use the information provided by grammar when they make lexical decisions, the prediction was that final-gemination non-words should be responded to more quickly than no-gemination non-words. However, the results of the experiment counter this – final-geminate forms were rejected more slowly than no-geminate forms. Similarly, the analysis developed in §2.1 implies that initial-geminate forms are more marked than no-geminate forms. The expectation is therefore that initial-geminate forms should be reacted to more quickly than no-geminate forms. This was also not confirmed in the experiments – no significant difference was found in the reaction times associated with initial-geminate forms and no-geminate forms. How should we interpret these negative and null results?

(33) **Average difference in response times between different types of non-words in *Experiment 2* of Berent *et al.* (2001b)**

Comparison	**Example**	**Difference score (*ms*)** [28]	***t***	***df***	***p***
Initial-gemination and no-gemination	Q-Q-S Q-S-M	15	–	–	> 0.05[29]
Initial-gemination and final-gemination	Q-Q-S Q-S-S	42	4.09	56	< 0.0001[30]
Final-gemination and no-gemination	Q-S-S Q-S-M	-27	2.48	56	< 0.009[31]

(34) **Lexical decision: Difference scores between the mean response times for the different token classes in the Berent *et al*. (2001b) data**



---

[28] These difference scores were computed by subtracting the average response time of the more marked token types from the average response times of the less marked token types. A positive difference score then indicates that the more marked tokens were rejected more quickly, and therefore counts as a success in terms of the hypothesis.

[29] Berent *et al*. do not report the actual *t*-statistic for this comparison. They do, however, state that there was no significant difference in mean response times for the initial-geminate forms and the no-geminate forms (Berent *et al*., 2001b:655).

[30] This *p*-value is for a one-tailed *t*-test, assuming a positive difference score. It therefore shows that initial-geminate forms were reacted to significantly more quickly than final-geminate forms.

[31] This *p*-value is for a one-tailed *t*-test, assuming a negative difference score. It therefore shows that final-geminate forms were reacted to significantly more slowly than no-geminate forms.

Berent *et al*. explain these unexpected results as follows: The more productive (phonological or morphological) processes are applied in the formation of a non-word, the more word-like it will be experienced to be by language users. An example from English might make the idea clearer: An English non-word such as “blick” will be experienced as less word-like than a non-word such as “blicked”, because a productive derivational process has applied in “blicked”. Berent *et al*. then assume that the subjects in their experiment interpreted all geminate forms (both initial and final) as derived via a productive gemination process from bi-radical roots. The fact that gemination was applied in the formation of these non-words lends a word-like feeling to them. This is then supposed to explain the slower than expected response times associated with geminate non-words.

However, there is a problem with this explanation of Berent *et al*. If application of process results in causing a non-word to be experienced as more word-like, then the effects of this should have been seen also in the word-likeness/well-formedness judgment experiments. For instance, we would then have expected that final-geminate forms would have been rated as more word-like than no-geminate forms. And this was not found. In fact, the opposite was found. I therefore suggest a different explanation for the unexpected response time results associated with geminate forms.

In no-geminate forms, the subjects consider only one possible derivation for the percept – from a tri-consonantal root to a tri-consonantal stem. However, both in initial and final-geminate forms, subjects consider two possible derivations – from a bi-consonantal root to a tri-consonantal stem, and from a tri-consonantal root to a tri-consonantal stem. (See the discussion in §2.1.7 about this.) Two derivational histories have to be considered for the geminate forms, and this is what slows down the reaction times associated with geminate forms.

Consider the comparison between initial-geminate and final-geminate forms. For both of these two kinds of tokens, subjects have to consider two derivational histories. There is therefore no processing difference between these two kinds of tokens. The only difference between them is in terms of their markedness/well-formedness. We do therefore expect to see the result predicted by the grammar – i.e. that initial-geminate forms should be rejected more quickly than final-geminate forms. And this is confirmed by the results of the experiment.

Now consider the comparison between the no-geminate forms and the two types of geminate forms (final and initial). According to the analysis developed in §2.1 these forms are related as follows in terms of their well-formedness: |No-gemination ™ Final-gemination ™ Initial-gemination|. Based on grammar alone, we would expect the following relation between the reaction times: RT(No-gemination) > RT(Final-gemination) > RT(Initial-gemination). Suppose that the additional processing associated with gemination adds about the same amount of time to both final and initial-geminate forms. We can then explain the results of the experiment as follows: (i) Based on grammar alone, we expect a smaller RT-difference between no-geminate forms and final-geminate forms, than between no-geminate forms and initial-geminate forms. (ii) The additional processing time added by gemination is longer than the expected RT-difference between no-gemination and final-geminate forms. The result is that final-geminate forms are rejected more slowly than no-geminate forms. (iii) However, the additional time added by the gemination is about as long as the expected RT-difference between no-geminate and initial-geminate forms. The result is that there is no appreciable difference in actual RT's between initial-geminate and no-geminate forms.

Under this interpretation we can also explain why the well-formedness judgment experiments did give the results predicted by the grammatical analysis. Also in the well-formedness judgment experiments subjects consider two possible derivational histories for the two kinds of geminate forms. However, these experiments were all self paced – i.e. subjects could take as much time as they needed to respond. The additional processing time in the geminate forms should therefore not influence the judgments of the subjects.

3. The processing of non-words of the form [sCvC] in English

In the previous section we saw evidence for the fact that language users make multi-level well-formedness distinctions. Hebrew speakers distinguish not only between possible and impossible words. They also distinguish between different possible words in terms of their well-formedness. This shows that we need a theory of grammar that can make multi-level well-formedness distinctions. In this section I will discuss a restriction on possible words in English that confirms this. I will discuss a set of experiments that show that English speakers distinguish between possible and impossible words in terms well-formedness. However, the results of the experiments show that in addition to this distinction English speakers also distinguish between different kinds of impossible words in terms of well-formedness. Together with the Hebrew results in §2, these results show that we need a theory of grammar that can make the following distinctions: (i) between possible and impossible words; (ii) within the set of possible words, between more and less well-formed possible words; (iii) also within the set of impossible words, between the more and less well-formed. We need a theory of grammar that can take any two forms and compare them for their relative well-formedness, whether they are possible or impossible words.

English places restrictions on what consonants can co-occur in the onset and coda position of a mono-syllabic word.[32] Restrictions of this kind were first noted by Fudge (1969), and include for example the following: (i) A mono-syllabic word cannot begin with [Cr] and end in [r] – i.e. **frer*, **krer*, etc. (ii) A mono-syllabic word cannot begin in [s] plus a nasal and end in nasal – i.e. **snam*, **smang*, etc. (iii) A mono-syllabic word cannot begin in [s] plus a nasal and end in [lC] – i.e. **snelk*, **smelk*, etc.

In this section I will investigate one of these restrictions in more detail. English does not allow words of the form [sCvC] where both C's are voiceless labial stops or voiceless velar stops. [sCvC]-words are allowed with two voiceless coronal stops – i.e. *state* is a word of English, but **spape* and **skake* are not even possible words (Browne, 1981, Clements and Keyser, 1983, Davis, 1982, 1984, 1988a, 1988b, 1989, 1991, Fudge, 1969, Lamontagne, 1993: Chapter 6).[33]

---

[32] These restrictions actually apply to syllables. A multi-syllabic word that has a syllable with any of these combinations will also not be a possible word of English. In the discussion here I will limit myself to mono-syllabic forms so that restrictions on possible syllables can also be treated as restrictions on possible words.

[33] This is also part of a larger restriction. For instance, a form with a voiced and voiceless labial is also not allowed, i.e. **spab*. However, forms with voiced and voiceless velars are at least marginally tolerated, cf. *skag*. Forms with voiceless stops and homorganic nasals are also generally not allowed, i.e. **skang*, **spim*.

The term *spam* is an exception. This term originated as a brand name for a kind of luncheon meat *SPAM*, and was later (inspired by a Monty Python sketch) extended to refer to unsolicited e-mail (http://en.wikipedia.org/wiki/Spamming#Etymology). Brand names, like other proper names, are often exempted from restrictions that apply to other words.

In the discussion here I will focus on this restriction only as it applies to the voiceless stops.

In the rest of this section I will discuss experimental evidence showing that this restriction has psychological reality for speakers of English. In particular, it influences the way in which they process non-words. This section is structured as follows: In §3.1 I develop an OT analysis of this restriction on possible words in English. The analysis establishes that [sCvC]-forms are related as follows in terms of their relative well-formedness: |sTvT ™ sKvK ™ sPvP|. In §3.2 I then discuss a set of well-formedness judgment and lexical decision experiments that I conducted to test the predictions of this analysis. The results of the experiments confirm that the restriction on [sCvC]-forms are psychologically real for speakers of English and that it influences the way in which they process non-words of this form.

3.1 Restrictions on [sCvC]-words

English does not tolerate words of the form [sCvC] where both C's are voiceless labial or velar stops. Forms with two voiceless coronal stops or with two heterorganic voiceless stops are tolerated.

(35) **Restrictions on [sCvC]-forms in English**

Allowed			**Not allowed**	
Two [t]'s		*state*	Two [p]'s	**spape*
Different C's	[k]/[t]	*skate, steak*	Two [k]'s	**skake*
	[p]/[t]	*spit, steep*		
	[p]/[k]	*skip, speak*		

The same restriction is found 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).[34] Twaddell found many words with the sequence [sTvT]. However, he found no words with the sequence [sKvK] or [sPvP]. The data in the table in (36) is extracted from Twaddell's 1940-paper.

(36) **Frequency count of [sCvC] sequences in German**[35]

st			**sk**			**sp**		
t	k	p	t	k	p	t	k	p
51	95	76	15	–	2	58	73	–

Like English, German tolerates [sCvC] words where the two C's are heterorganic voiceless stops, or two voiceless coronal stops, but not where C's are voiceless labial or velar stops. The examples in (37) illustrate this point.[36]

Afrikaans is different from German and English. Like German and English Afrikaans does not tolerate [sCvC] words where both C's are voiceless labial stops. However, unlike German and English, forms with two voiceless velar stops are well-formed in Afrikaans. This is illustrated by the examples in (38).[37]

---

[34] Twaddell did not take syllabic boundaries into consideration. His counts would therefore include also forms with syllable boundaries after the vowel, i.e. [sCv.Cv…]. For the purposes of the discussion here, what is important is that syllables of the form [sCvC] are included in his counts. Another caveat is in order – Twaddell did not include the unstressed syllables in his counts. His counts alone are therefore not enough to substantiate the claim that German has the same restriction as English. The intuition that [sPvP] and [sKvK] sequences are ill-formed in German was confirmed by consultation with native speakers of German. See footnote 38 below.

[35] Since [s] preceding a consonant is often pronounced as [ʃ] in German, Twaddell included both [sC] and [ʃC] sequences in these counts.

[36] I am indebted to my German consultants, Florian Schwarz and Tanja Vignjevic, for supplying these examples. They also confirmed the intuition that [sPvP] and [sKvK] forms are ill-formed in German.

[37] Afrikaans is my native language. I therefore base my statements about Afrikaans on my own intuitions.

(37) **Restrictions on [sCvC]-forms in German**

Allowed				**Not allowed**	
Two [t]’s		*Staat*	“state”	Two [p]’s	**Spaap*
		Stadt	“city”		**Spep*
Different C’s	[k]/[t]	*Stock*	“stick”	Two [k]’s	**Skaak*
		Skat	“card game”		**Skek*
	[p]/[t]	*Stop*	“stop”		
		Spott	“ridicule”		
	[p]/[k]	*Skip*	“leader of team in curling”		
		Speck	“bacon”		

(38) **Restrictions on [sCvC]-forms in Afrikaans**

Allowed				**Not allowed**	
Two [t]’s		*staat*	“state”	Two [p]’s	**spaap*
		stad[38]	“city”		**spep*
Two [k]’s		*skok*	“shock”		
		skaak	“chess”		
Different C’s	[k]/[t]	*stok*	“stick”		
		skat	“treasure”		
	[p]/[t]	*stop*	“stop”		
		spot	“ridicule”		
	[p]/[k]	*skip*	“ship”		
		spek	“bacon”		

---

[38] Afrikaans has coda devoicing. This word therefore ends on a [t] in pronunciation.

[sTvT]-forms are possible words in English, German and Afrikaans, and [sPvP]-forms are not possible words in any of these languages. [sKvK]-forms take the intermediate position. These forms are possible words in Afrikaans, but not in English or German. Based on this, I propose the following markedness/harmony scale: |sTvT ™ sKvK ™ sPvP|.

In the rest of this section I develop an OT analysis of this restriction on possible words in English. In this analysis I will assume the markedness/harmony scale |sTvT ™ sKvK ™ sPvP|. Even though neither [sKvK]-forms nor [sPvP]-forms are possible words in English, I will assume that [sKvK]-forms are more well-formed than [sPvP]-forms. Section §3.1.1 contains a basic OT analysis. Section §3.1.2 then contains an extensive motivation of the analysis, showing that it is well founded both cross-linguistically and theory internally. Section §3.1.3 considers the predictions that follows from the analysis with regard well-formedness judgments and lexical decision reaction times.

3.1.1 A basic OT analysis

Following a tradition that originated with Prince and Smolensky (1993), I assume that there is a markedness constraint against every element on a harmony scale. The markedness constraints are ranked in an order opposite to the elements on the harmony scale, so that the element that occupies the lowest slot on the harmony scale violates the highest ranked markedness constraint. This is illustrated in (39) for the harmony scale associated with [sCvC]-forms.[39]

---

[39] De Lacy (2002) has a different view. Rather that assuming a fixed ranking between markedness constraints that refer to a harmony scale, he argues for constraints that can freely rerank but that are in a stringency relation (Prince, 1998). Under this interpretation of the relationship between harmony scales and markedness constraints [sTvT] will violate only *sTvT, [sKvK] will violate *sKvK and

(39) **From harmony scales to markedness constraint hierarchy**

Harmony scale: |sTvT ™ sKvK ™ sPvP|

Markedness constraint hierarchy: ||*sPvP o *sKvK o *sTvT||

(Where ***sC$_i$vC$_i$** = Do not allow the sequence [sC$_i$vC$_i$] within a single syllable.)

English tolerates violation of the constraint *sTvT. This implies that all faithfulness constraints that could be violated in order to avoid violation of *sTvT should outrank *sTvT. On the other hand, English does not tolerate violation of the constraints *sKvK or *sPvP. These two constraints must therefore outrank at least some faithfulness constraint that could be violated in order to avoid violation of *sKvK or *sPvP. Since there is no active process in English that shows how English would avoid violation of *sPvP and *sKvK, it is not possible to decide definitively what the relevant faithfulness constraint is. Violation of these constraints can be avoided in many different ways. In (40) some possibilities are listed, together with the faithfulness constraint that each repair strategy would violate.

---

*sTvT, and [sPvP] will violate *sPvP, *sKvK, and *sTvT. Although the constraints can freely rerank, the harmonic ordering |sTvT ™ sKvK ™ sPvP| will still hold. Whether we use markedness constraints in a fixed ranking or stringency related constraints, the same harmonic ordering between the [sCvC]-forms hold. The main claims that I make in this section is that this is the harmonic ordering that grammar imposes on [sCvC]-tokens and that this ordering influences the manner in which English listeners process [sCvC]-tokens. Since this ordering is imposed both by constraints in a fixed ranking and by stringency related constraints, it does not matter for my purposes which of the two analyses I use.

Gouskova (2003) uses constraints in a fixed ranking, but shows that there is no constraint against the most harmonic element on a harmony scale. However, even under Gouskova’s interpretation of harmony scales there will be a constraint against [sTvT]-forms. The reason is that [sTvT] is not the least marked form. Forms with heterorganic voiceless stops probably occur at the least marked end of the harmonic ordering, i.e. |sTvK ™ sTvT ™ sKvK ™ sPvP|. The existence of the constraint *sTvT is then not contrary to Gouskova’s interpretation of harmony scales.

(40) **Avoiding violation of *sCvC**

Input	**Repair**	**Example output**	**Faithfulness constraint violated**
/sKvK/	Deletion	[sKv], [KvK]	MAX
	Epenthesis	[sKv.Kv]	DEP
	Place change	[sKvT], [sPvK]	IDENT[place]

In the rest of the discussion I will assume that the relevant faithfulness constraint is IDENT[place]. However, this is an arbitrary choice – any of the other constraints in (40) would have done equally well. If we rank IDENT[place] between *sKvK and *sTvT, we can account for the fact that [sTvT] is a possible word of English, while [sKvK] and [sPvP] are not. This is shown in the tableaux in (41).

(41) **a.** **[sTvT] possible word**

/sTvT/ → [sTvT]

/sTvT/	*sPvP	*sKvK	IDENT[place]	*sTvT
☞ sTvT				*
sTvK			*!	

b. **[sKvK] not possible word**

/sKvK/ → [sKvT]

/sKvK/	*sPvP	*sKvK	IDENT[place]	*sTvT
sKvK		*!		
☞ sKvT			*	

c. **[sPvP] not possible word**

/sPvP/ → [sPvT]

/sPvP/	*sPvP	*sKvK	IDENT[place]	*sTvT
sPvP	*!			
☞ sPvT			*	

These tableaux show that an /sTvT/-input will be mapped faithfully onto itself, confirming that [sTvT] is a possible word. However, for an /sKvK/-input, an unfaithful candidate will be selected over the faithful [sKvK]. This shows that [sKvK] is not a possible word of English. The same is true for an /sPvP/-input.

3.1.1.1 The critical cut-off

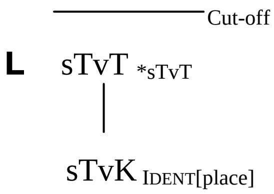
We need to locate the critical cut-off point in this mini-hierarchy for English. (For a discussion of the critical cut-off, refer to Chapter 1 §2.2.3, Chapter 3 §2.3.) Due to the lack of evidence of variation in the pronunciation [sCvC]-forms in English, I make the conservative assumption that the critical cut-off is located at the bottom of the mini-hierarchy developed above (see Chapter 4 §3.1 on this assumption). All candidates will therefore violate at least some constraint ranked higher than the cut-off. In these situations, only the best candidate is accessible as potential output. This assures that no variation will be observed in the pronunciation of [sCvC]-forms in English. This is illustrated for an /sTvT/-input in (42). However, the same point can be made for /sPvP/-inputs and /sKvK/-inputs. As always, a thick vertical line is used to indicate the place of the critical cut-off in the hierarchy, and a graphic representation of the ordering imposed by EVAL on candidate set is given next to the tableau.

(42) **Adding in the critical cut-off**

/sTvT/ → [sTvT]

/sTvT/	*sPvP	*sKvK	IDENT[place]	*sTvT
1 sTvT				*
2 sTvK			*!	

Output of EVAL



3.1.1.2 Interim summary

The table in (43) contains a summary of the constraints and rankings argued for thus for. The first column contains the rankings, the second column a short motivation for the ranking, and the third column an indication of where in the preceding discussion the ranking was discussed.

(43) **Summary of the constraints and rankings for English**

Ranking	Motivation	Where motivated
*sTvT o {*sKvK, *sPvP}	[sTvT] possible word [sKvK], [sPvP] not	(35)
*sKvK o *sPvP	[sKvK] possible word in Afrikaans, [sPvP] not	(38)
IDENT[place] o *sTvT	[sTvT] possible word	(41)
{*sKvK, sPvP} o IDENT[place]	[sKvK], [sPvP] not possible words	(41)
*sTvT o cut-off	No variation	(42)

3.1.2 Motivation of analysis

The analysis presented above depends on two assumptions: (i) the existence of *sCvC-constraints; and (ii) that these constraints are ranked ||*sPvP o *sKvK o *sTvT||. In this section I discuss the evidence for these assumptions. The aim of this section is to show that these assumptions are well founded both cross-linguistically and theory internally. This section is structured as follows: In §3.1.2.1 I discuss the evidence in favor of the |coronal ™ velar ™ labial| place harmony scale in English. This serves as evidence for the ranking between the *sCvC-constraints. Section §3.1.2.2 motivates the existence of OCP-constraints against multiple occurrences of [t], [k] or [p] within a single syllable. Section §3.1.2.3 then presents evidence that [s+stop]-structures are marked. Finally, in §3.1.2.4 I show that the *sCvC-constraints are just the local conjunction of the OCP-constraints with a constraint against [s+stop]-structures.

3.1.2.1 The place harmony scale in English

In the analysis above I assumed the ranking ||*sPvP o *sKvK o *sTvT||. Implicit in this ranking is the idea that labial place is more marked than velar place, which is again more marked than coronal place. This is expressed formally in the place harmony scale in (44).

(44) **Place harmony scale**

|coronal ™ velar ™ labial|

In this section I will provide evidence that (44) is the correct representation of the place harmony scale for English. I begin by presenting evidence that coronals are the most harmonic, establishing that |coronal ™ {velar, labial}| holds of English. I then discuss the evidence that velars are more harmonic than labials, establishing that |velar ™ labial|.

It is generally accepted that coronal place is universally less marked than velar and labial place (de Lacy, 2002, Jakobson, 1968, Lombardi, 2001, Paradis and Prunet, 1990, 1991, Prince, 1997, 1998). On these grounds it can then be expected that the ordering |coronal ™ {labial, velar}| should hold of English too. However, there is also more direct evidence from phonotactic restrictions in English that coronals are less marked than velars and labials.

There are certain sequences that are allowed with coronal consonants, but not with velar or labial consonants. Of course, the [sCvC]-forms that are the focus of the discussion represent one such example. But there are more. English allows word final clusters of up to four consonants. The first and second consonant can be of any place of articulation. However, the third and fourth position can only be filled by coronals (Fudge, 1969). This shows that the restrictions on the distribution of coronals are less strict than the restrictions on the distribution of velars and labials.

(45) **Word final consonants in English**

	Position 1	**Position 2**	**Position 3**	**Position 4**
Coronal	*cat*	*lint*	*pasts*	*sixths*
Velar	*kick*	*crank*	–	–
Labial	*dam*	*lamp*	–	–

In order to determine the harmonic ordering between velars and labials, we can also use phonotactic restrictions. There are certain sequences that are allowed with coronals and velars but not with labials. For instance, sequences of the form [sCvNC] are allowed with coronal and velars but not with labials – where both C's are voiceless homorganic stops and N is a nasal homorganic to the stops. This shows that labials are more restricted in their distribution than either velars or coronals.

(46) **Words of the form [sCvNC]**

Coronals	*stint*	*stunt*
Velars	*skink*	*skunk*
Labials	**spimp*	**spump*

The co-occurrence patterns of homorganic consonants in the English lexicon also suggests that labials are more marked than velars, which are again more marked than coronals. Using a dictionary with approximately 20,000 English words, Berkley (1994a, 1994b, 2000) calculated the number of words with two labials, two velars, and two coronals. She then calculated the number of such words that would have been expected had the consonants been allowed to combine freely. The observed/expected ratio for each of these three places of articulation can then be calculated. An observed/expected ratio of smaller than 1 means that fewer words of that structure occur than what are expected had the consonants been allowed to combine freely.[40] Berkley found that words with two labials are more under represented than words with two velars, and that words with two velars are again more under represented than words with two coronals. This suggests that labials are subject to stronger occurrence restrictions that velars, which again are subject to stronger restrictions than coronals. The table in (47) is based on data extracted from Berkley (2000:23-30), and is representative of the patterns that Berkley found. This table contains the statistics about words with two labials, velars or coronals separated by one or two segments.

(47) **Co-occurrence patterns of two labials, velars or coronals separated by one or two segments**

	Example	**Observed**	**Expected**	**Observed/Expected**
Labials	*pop*, *pulp*	118	207.9	0.57
Velars	*cock*, *crack*	81	113	0.71
Coronal	*ten*, *tact*	1148	1271.5	0.90

---

[40] In this she follows a method first used for Arabic by *inter alia* Pierrehumbert (1993) and Frisch *et al.* (2004).

These data show that all three kinds of words are under-represented (the observed/expected ratios for all three classes are smaller than 1). However, words with two labials are more under-represented than words with two velars, which is again more under-represented than words with two coronals.

In summary: (i) English restricts the occurrence of labials and velars more than the occurrence of coronals. This supports the harmonic ordering |coronal ™ {velar, labia}|. (ii) Similarly, English restricts the occurrence of labials more than velars, supporting this harmonic ordering |velar ™ labial|.

3.1.2.2 Constraints against the co-occurrence of homorganic voiceless stops

In the previous section we have seen evidence for the fact that English restricts the co-occurrence of homorganic voiceless stops in certain configurations. In order to capture this, I suggest the existence of OCP-constraints that penalize the co-occurrence of identical segments within a certain domain. In (48a) I give the general schema for these constraints, in (48b) the specific instantiations that apply to voiceless stops, and in (48c) the ranking that holds between these constraints in English.

(48) **OCP constraints**

a. **General** $*[\alpha \ldots \alpha]_\delta$: Do not allow two [α]'s in domain δ.

b. **Coronal** $*[t \ldots t]_\sigma$: Do not allow two [t]'s in one syllable.
Velar $*[k \ldots k]_\sigma$: Do not allow two [k]'s in one syllable.
Labial $*[p \ldots p]_\sigma$: Do not allow two [p]'s in one syllable.

c. **Ranking** $\|*[p \ldots p]_\sigma \; \mathrm{o} \; *[k \ldots k]_\sigma \; \mathrm{o} \; *[t \ldots t]_\sigma\|$

In the formulation of these constraints I assume that the domain over which they apply is the syllable. This domain can be defined differently. For instance, it might be that the domain should be morphologically defined as a word or a morpheme (Tessier, 2003, 2004). In the experiments that I will report in §3.2 below all tokens were of the form [sCvC]. These tokens were therefore all mono-syllabic, and there was no reason for subjects to assume that the tokens were not also mono-morphemic. For these specific tokens it would not matter whether the domain is defined as the syllable, the word or the morpheme. The constraint would apply to these forms under any of these definitions. For this reason I will not consider the implications of the different possible definitions of the domain here.[41]

In the OT literature OCP-constraints are often defined as the local self-conjunction of markedness constraints (Alderete, 1996, 1997, Itô and Mester, 1998, Smolensky, 1995). The constraints in (48b) would then be stated as the local self-conjunction of constraints against a singleton coronal, velar or labial stop. For instance, in stead of $*[p \ldots p]_{\sigma}$ we would have $*[p]_{\sigma}^{2}$ (do not violate *[p] twice in the same syllable).

I opt not to adopt this definition of OCP-constraints. Viewing an OCP-constraint $*[\alpha \ldots \alpha]_{\delta}$ as the local self-conjunction of the constraint $*[\alpha]$ presupposes the existence of the constraint against a single occurrence of $[\alpha]$. If we redefined the constraints in (48b) in terms of local self-conjunction, then we are presupposing of the existence of the constraints *[p], *[k] and *[t]. The markedness of [k] and [p] can be motivated by referring to the universal harmony scale |coronal ™ {velar, labial}| (see §3.1.2.1). Since

---

[41] See also the discussion on domains in §3.1.2.4 below.

velars and labials are marked relative to coronals, [k] and [p] are marked relative to [t]. However, coronal appears as the most harmonic member of this scale, and Gouskova (2003) has recently shown that markedness constraints against the most harmonic member of a harmony scale results in incorrect typological predictions. The existence of *[t] is therefore questionable.

The intuition behind the formulation of the constraints in (48) can be stated as follows: even if a single instance of some structure is not marked, multiple occurrences of that structure in some localized domain might still be marked. Even though [t] is not individually marked, two occurrences of [t] within a single syllable is marked. This can be expressed in terms of a harmony scale: |α_δ ™ (α … α)$_\delta$|. This scale should be read as follows: one occurrence of the structure α within domain δ is more harmonic than two occurrences of α within domain δ. In Gouskova’s interpretation of such harmony scales there will not be a constraint against the least marked member of the scale (no constraint against α), but there will be a constraint against all other members of the scale (against two occurrences of α in domain δ).

The only difference between the three constraints in (48b) is the place of articulation to which each refers. I therefore assume that their ranking reflects the place harmony scale |coronal ™ velar ™ labial| motivated in the previous section (§3.1.2.1). The constraints are ranked so that the constraint that refers to most harmonic place of articulation (coronal) is ranked lowest.

English tolerates violation of all three of these constraints. There are many words with two [t]'s, [k]'s or [p]'s in the same syllable (*tot*, *cake*, *pop*). This implies that these OCP-constraints should rank below all faithfulness that could be violated in order to avoid violating the OCP-constraints.

3.1.2.3 On the markedness of [s+stop]-structures

In §3.1.2.4 I will show that each of the *sCvC-constraints is formed by locally conjoining the OCP-constraints from (48b) with a constraint against [s+stop]-structures. The existence of the OCP-constraints has been motivated in the previous section. However, I still need to show that there is a constraint against [s+stop]-structures. In (49) I state this constraint, and the rest of this section is then dedicated to motivating the existence of this constraint.

(49) ***[s+stop]**

Do not allow a tautosyllabic [s+stop]-sequence.

There is general acceptance in the literature that [s+stop]-structures are marked. However, there is no agreement on the reason for their markedness. The approaches in the literature can be classified into two broad groups: (i) *Cluster approach*. Under this approach these sequences are interpreted as consonant clusters. The markedness of the sequences (at least in word initial position) then results from the fact that they violate the sonority sequencing principle (Selkirk, 1982a). (ii) *Complex segment approach*. Under this approach an [s+stop]-structure is interpreted as a single complex segment rather than a sequence of two separate segments. The markedness of the structure then follows from the fact that it is complex segment (Sagey, 1990).

In order to explain the results of the experiments that I report in §3.2 the reason for the markedness of the [s+stop]-structures is not relevant. What is relevant is that these structures are marked, and that there is a constraint against them. I will therefore not choose between the two accounts for the markedness of these sequences. The rest of this section consists of discussion of a representative sample of the arguments in favor of both views on the markedness of [s+stop]-structures. The purpose of this discussion is to show that these structures are marked, whether they are viewed as consonant clusters or as complex segments.[42]

Cluster approach.[43] The first argument in favor of interpreting [s+stop]-structures as consonant clusters rests on English stress placement rules and comes from Hayes (1980, 1982, 1985). English regularly stresses the ante-penultimate syllable. However, when the penultimate syllable is heavy, stress is attracted from the ante-penult to the heavy penult. Hayes points out that words with an [s+stop]-structure preceding the final vowel are stressed as if the penultimate syllable is closed – i.e. stress falls on the penult. This can be explained if we assume that the [s] syllabifies into the coda of the penultimate syllable and the [stop] into the onset of the ultimate syllable. However, if a syllable boundary intervenes between the [s] and the [stop], then this structure should be interpreted as a sequence of two consonants. The examples in (50) illustrate this point.

---

[42] In fact, it is quite possible that both approaches are correct. It is possible that [s+stop]-structures are clusters in some languages and complex segments in others. It is even possible within one language that not all [s+stop]-structures have the same structural representation. See Fleischhacker (2001, 2002) for arguments to this effect.

[43] For examples of proponents of this approach, see *inter alia* Davis (1984), Hayes (1985:1480149), Hockett (1955:152-153), Gouskova (2001), Kahn (1980:42-43).

(50) **Stress placement in English**

Penult light = stressed ante-penult	**Penult heavy = stressed penult**
Cánada	*agénda*
sýllable	*patáto*
prínciple	*amálgam*

[s+stop] before final vowel acts like heavy penult words

seméster	**sémester*
Damáscus	**Dámascus*
Gilléspie	**Gíllespie*[44]

Closely related to this argument based on stress placement, is the argument based on the phenomenon of "pre-stress destressing" in English. This argument is also due to Hayes (1980, 1982, 1985). In pre-stress position English regularly de-stresses and consequently neutralizes vowels. However, the neutralization is blocked if the pre-stress vowel occurs in a closed syllable. Hayes points out that a pre-stress vowel followed by an [s+stop]-structure is also spared from neutralization. This can be explained if we assume that the [s] syllabifies into the coda of the pre-stress syllable, and the [stop] into the onset of the stressed syllable. Again, the intervening of a syllable boundary between the two parts of the structure is easier to explain if we assume that these structures are in fact consonant clusters. The examples in (51) illustrate this point.

---

[44] But see Davis (1982) and Lamontagne (1993) for counter examples such as *pédestal*, **pedéstal* and *áncestor*, **ancéstor*. According to Davis there are just about an equal number of words like *pédestal* and words like *seméster*.

(51) **Pre-stress destressing and neutralization in English**

Pre-stress open = neutralization	**Pre-stress closed = no neutralization**
s[ə]*méster*	*d*[ɔ]*ctórial*
[ə]*mérica*	*c*[æ]*ntéén*
C[ə]*nnécticut*	*sh*[æ]*mpóó*

Pre-stress vowels followed by [s+stop]-sequences resist neutralization

pl[æ]*stícity*	**pl*[ə]*stícity*
m[æ]*scára*	**m*[ə]*scára*
m[æ]*stítis*	**m*[ə]*stítis*[45]

More evidence for the cluster approach comes from trends in syllabification. Treiman (1983), Treiman and Danis (1988) and Treiman *et al.* (1992) conducted a series of experiments in which they investigated how English speakers syllabify intervocalic consonant clusters. In general they found that, if two inter-vocalic consonants would form a licit onset, English speakers prefer to syllabify the consonants together into the onset of the second syllable – i.e. when presented with a token such [nəprim] speakers more often syllabified as [nə.prim] than as [nəp.rim]. However, they found the opposite for [s+stop]-structures. Speakers more often placed a syllable boundary between the [s] and the [stop] – i.e. when presented with a token such as [nəspim], speakers more often syllabified as [nes.pim] than as [ne.spim]. This suggests that the [s] and [stop] parts of these structures

---

[45] Again see Davis (1982) and Lamontagne (1993) for counter examples such as [ə]*stónish* and *n*[ə]*stálgia*.

are not very closely connected, and therefore than they are clusters rather than complex segments.

There is also evidence from languages other than English that [s+stop]-structures act like clusters rather than single segments. I will discuss an example from Italian here. This example is due to Kaye, Lowenstamm and Vergnaud (1990). Italian has a process that lengthens stressed vowels in open syllables. Stressed vowels in closed syllables, however, do not undergo this process. Stressed vowels followed by an [s+stop]-structure are also immune to the lengthening process. This can be understood if we assume that the [s] is syllabified into the coda of the stressed syllable and the [stop] into the onset of the following syllable. Again, the occurrence of a syllable boundary between the [s] and the [stop] can be explained more easily by assuming that these structures are consonant clusters rather than single complex segments.

(52) **Stressed syllable lengthening in Italian**

Open = lengthening		**Closed = no lengthening**	
_f_[á:]_to_	‘fate’	_f_[á]_tto_, *_f_[á:]_tto_	‘fact’
_m_[é:]_ro_	‘pure’	_m_[á]_nto_, *_m_[á:]_nto_	‘coat’

Pre-[s+stop] = no lengthening	
_p_[á]_sta_, *_p_[á:]_sta_	‘pasta’
_p_[é]_sta_, *_p_[é:]_sta_	‘trail’

There is therefore ample evidence, both in English and cross-linguistically, that [s+stop]-structures are (sometimes) treated as consonant clusters. If we assume this structure for word initial [s+stop]-structures in English, then the markedness of these structures can be explained easily. The “sonority sequencing principle” dictates that the sonority of consonants in the onset of a syllable cannot fall towards the nucleus (Selkirk, 1982a). The fricative [s] is higher is sonority than a stop (Parker, 2002). Consequently, an [s+stop] onset cluster violates the sonority sequencing principle.

Complex segment approach.[46] The first two arguments in favor of the complex segment approach involve syllable structure in English and come from Selkirk (Selkirk, 1982b). If we assume that [s+stop]-structures are clusters, then these clusters represent the only onsets in English that violate the sonority sequencing principle. All other onset clusters rise in sonority towards the nucleus, but [s+stop]-clusters fall in sonority. If [s+stop]-structures are interpreted rather as complex segments, then there are no exceptions to the sonority sequencing principle for onsets in English.

Closely related to this is the argument about the number of consonants that can occur in onset position in English. If we interpret [s+stop]-structures as complex segments then onsets in English are maximally two segments long. However, if [s+stop]-structures are interpreted as consonant clusters, then English does allow tri-consonantal onset clusters (cf. *<u>str</u>eet*, *<u>spl</u>ash*, *<u>sq</u>ueak*). All tri-consonantal onset clusters then have the structure [s + stop + glide/liquid]. Interpreting [s+stop]-structures as clusters therefore results in a complication to the syllable structure of English (tri-consonantal onsets are now allowed), and in an inexplicable idiosyncrasy of the tri-consonantal clusters (why [s + stop + glide/liquid] is the only observed structure). Both of these can be avoided by assuming that [s+stop]-structures are complex segments rather than consonant clusters.

---

[46] For examples of proponents of this approach, see *inter alia* Broselow (1991), Firth (1935, 1935-1937:543), Fudge (1969), Fujimura and Lovins (1978:112), Lamontagne (1993: Chapter 6), Selkirk (1982b), Spang-Hanssen (1959:158-161).

Also for the complex segment approach, evidence from languages other than English is available. I will discuss two examples here. First, about alliteration in Late-Middle and Early-Modern Irish verse and Old Germanic verse. These traditions made extensive use of alliteration as a poetic device. The rules determining what sounds can alliterate with what sounds were complicated. However, the rules in (53) capture the basic patterns (Kurylowicz, 1971, Meyer, 1909, Murphy, 1961).

(53) **Rules of alliteration in Irish and Germanic verse**

- (i) Any vowel alliterated with any other vowel.
- (ii) A consonant alliterated only with itself.
- (iii) But [s+stop]-initial words could not alliterate with [s]-initial words. [st-] alliterated only with itself, and similarly for [sk-] and [sp-].

The peculiar behavior of [s+stop]-initial words can be explained more easily if it is assumed that the [s+stop]-structures formed single complex segments.

Another example of a language where [s+stop]-structures are treated as complex segments, is Egyptian Arabic. Egyptian Arabic allows no onset clusters. When Egyptian Arabic borrows words from English that start on a consonant clusters, vowels are inserted in order to break up the word initial consonant cluster. However, when English words that start on [s+stop] are borrowed into Egyptian Arabic, this does not happen. The examples in (54) are from Broselow (1991).

Unlike other consonant sequences, the sequence [s+stop] cannot be broken up by an epenthetic vowel. This is more easily understandable if we assume that these sequences are interpreted as single complex segments rather than consonant clusters.

(54) **English borrowings into Egyptian Arabic**

#CC → #CvC		**#s+stop → # ʔ i s+stop**	
plastic	[bilastik]	*ski*	[ʔiski]
Fred	[fired]	*study*	[ʔistadi]
sweater	[siwetar]	*spring*	[ʔispriŋ]
slide	[silaid]	*street*	[ʔistiri:t]

We therefore have evidence both from within English and cross-linguistically that [s+stop]-structures are (at least sometimes) treated as single complex segments. If we assume this interpretation for [s+stop]-structures, then they are marked simply by virtue of being complex segments. Complex segments have either multiple places of articulation or multiple manners of articulation (or both), and are as such marked relative to ordinary simplex segments. It is also true that complex segments are relative scarce in the segmental inventories of the world's languages (Ladefoged and Maddieson, 1996, Maddieson, 1984). This also serves as motivation for their markedness.

Whether we interpret word initial [s+stop]-structures in English as consonant clusters or as complex segments, it is clear that they are marked. If they are interpreted as consonant clusters, then they are marked by virtue of violating the sonority sequencing principle. If they are interpreted as complex segments, then they are marked *per se*. There must therefore exist some constraint that is violated by these structures. I use *[s+stop] from (49) to represent this constraint.

3.1.2.4 *sCvC-constraints as local conjunctions

In this section I will argue that each of the *sCvC-constraints is formed via the local conjunction (Smolensky, 1995) of OCP-constraints in (48b) with the constraint *[s+stop] from (49). Local conjunction is a well motivated algorithm for constructing complex constraints from simple constraints. The constraints that need to be conjoined to form the *sCvC-constraints are also well motivated (see discussion in the previous two sections). Both the method of constructing the *sCvC-constraints and the components that go into their construction are therefore well-motivated.

We have now motivated the existence of the OCP-constraints ||*[p … p]$_\sigma$ o *[k … k]$_\sigma$ o *[t … t]$_\sigma$|| and of the constraint *[s+stop]. English tolerates violation of each of these constraints individually, and even simultaneous violation of *[t … t]$_\sigma$ and *[s+stop]. What English does not tolerate is simultaneous violation of *[s+stop] and *[p … p]$_\sigma$ or *[k … k]$_\sigma$. A few examples are given in (55).

This is not an unknown phenomenon. There are countless examples of languages that tolerate violation of individual markedness constraints, but not the violation of certain combinations of markedness constraints in some local context. For instance, Itô and Mester (1997, 1998) analyze coda devoicing in German in this manner. German (and also Dutch, Afrikaans, Russian, etc.) tolerates both coda consonants and voiced obstruents separately, but not voiced obstruents in coda position. This generalization can be stated as follows in terms of markedness constraints: German tolerates violation of the constraint against having consonants in coda position (NoCoda) and of the constraint against voiced obstruents (*VoicedObs). What German does not tolerate, is simultaneous violation of both of these constraints in one syllabic position – i.e. German does not tolerate violation of NoCoda and *VoicedObs in the same coda. This then explains why a word such as *Rades* [ra:.dəs] 'wheels' is well-formed in German, while a word such as *[ra:d] is not. The form [ra:.dəs] violates *VoicedObs because of the [d]. It also violates NoCoda because its final syllable is closed. However, these two constraints are not violated in the same coda. The form *[ra:d] also violates both *VoicedObs and NoCoda. This form, however, is ill-formed because these two constraints are both violated in the same coda position.

(55) **Violation of OCP-constraints and *[s+stop] in English**

a. **Tolerated**

Example	**Constraints violated**
sting	*[s+stop]
skate	*[s+stop]
speak	*[s+stop]
tot	*[t … t]$_\sigma$
cock	*[k … k]$_\sigma$
pop	*[p … p]$_\sigma$
state	*[s+stop] & *[t … t]$_\sigma$

b. **Not tolerated**

**skak*	*[s+stop] & *[k … k]$_\sigma$
**spap*	*[s+stop] & *[p … p]$_\sigma$

The restriction that English places on [sCvC]-words is therefore not unusual. Both Alderete (1996, 1997) and Itô and Mester (1998) suggested that restrictions like these can be analyzed via local conjunction of markedness constraints. The locally conjoined constraint is then violated only if all the individual constraints that are conjoined to form the conjoined constraint are violated simultaneously. The three *sCvC-constraints can now be shown to be formed from the local conjunction of an OCP-constraint with the constraint against [s+stop]-structures. These constraints are stated in these terms in (56).

(56) ***sCvC-constraints as local conjunction**

sTvT = [[s+stop] & *[t … t]$_\sigma$]$_\sigma$

sKvK = [[s+stop] & *[k … k]$_\sigma$]$_\sigma$

sPvP = [[s+stop] & *[p … p]$_\sigma$]$_\sigma$

Itô and Mester (1998) formulate a principle that they call "ranking preservation". This principle requires the following: Let LC$_1$ and LC$_2$ be two constraints formed via local conjunction, and let C$_1$ be one of the conjuncts of LC$_1$ and C$_2$ one of the conjuncts of LC$_2$. If ||C$_1$ o C$_2$||, then ||LC$_1$ o LC$_2$||. I have argued in (48c) above that the OCP-constraints are ranked as follows for English: ||*[p … p]$_\sigma$ o *[k … k]$_\sigma$ o *[t … t]$_\sigma$||. *[p … p]$_\sigma$ is one of the conjuncts of *sPvP, *[k … k]$_\sigma$ of *sKvK, and *[t … t]$_\sigma$ of *sTvT. If we assume the principle of ranking preservation, it then follows that the three *sCvC constraints are indeed ranked as I assumed in §3.1.1 – i.e. ||*sPvP o *sKvK o *sTvT||.

Locally conjoined constraints are *locally* conjoined – this means that they are evaluated in some local domain. It is not the case that the locally conjoined constraint is always violated when all of its conjuncts are violated, but only when all of its conjuncts are violated within some local domain. In (56) I make the assumption that the relevant domain is the syllable. As with the OCP-constraints this domain could be stated differently (see the discussion in §3.1.2.2). The domain could also be defined morphologically as the word or the morpheme. In the experiments that I will discus below in §3.2 the tokens were of the form [sCvC]. All tokens were therefore monosyllabic, and there was no reason for subjects to assume that the tokens were poly-morphemic. Irrespective of whether we define the domain of the *sCvC-constraints as the syllable, the morpheme or the word, all of the tokens in the experiment would violate one of the *sCvC-constraints. For our purposes it is therefore not crucial which of these options we choose. I am making the arbitrary choice of assuming that the syllable counts as the domain.

I have shown above: (i) that the *sCvC-constraints are formed by the local conjunction of constraints that are all individually motivated both in English and cross-linguistically, and (ii) that the ranking ||*sPvP o *sKvK o *sTvT|| is based on the place harmony scale |coronal ™ velar ™ labial|. The basic OT analysis developed in §3.1.1 is therefore well-motivated. In the next section, I consider the predictions that follow from this analysis with regard relative well-formedness judgments and lexical decision reaction times.

3.1.3 The processing of non-words of the form [sCvC]

In §3.1.1 I developed the following mini-grammar for English: ||*sPvP o *sKvK o IDENT[place] o *sTvT o Cut-off||. Assuming that this analysis is correct, what predictions does it make about how English speakers will process non-words of the form [sCvC]? Recall the basic assumption about how grammar influences well-formedness judgments and lexical decision reaction times (§1.1): (i) The more well-formed a token is according to the grammar of the language, the more well-formed it will be judged to be by language users. (ii) The more well-formed a non-word is according the grammar of the language, the more seriously language users will consider it as a possible word, and the longer they will take to reject it as a non-word.

In the experiments that I conducted, I included non-words of the form [sTvT], [sKvK] and [sPvP]. What underlying representation would the subjects have assumed for these tokens? Consider an [sPvP]-token as an example. If subjects assumed an underlying representation identical the token, then they will assume the mapping /sPvP/ → [sPvP]. This mapping will violate at least *sPvP. If they assumed any underlying representation other than /sPvP/, the map /input/ → [sPvP] will violate in addition to *sPvP also some faithfulness constraint. Say that the subjects assumed the underlying representation /sPvT/. The mapping /sPvT/ → [sPvP] will then violate both *sPvP and IDENT[place]. The assumption that the underlying representation is identical to the perceived token therefore results in the more harmonic mapping. For this reason, I will assume that the subjects assumed an /sPvP/-input for all [sPvP]-tokens. The same argument can be made for [sKvK]-tokens and [sTvT]-tokens. Under these assumptions, these three kinds of non-words can be compared as shown in (57) in the mini-grammar developed above. Since all three tokens are assumed to be faithful candidates, the faithfulness constraint IDENT[place] is not violated.

(57) **Comparing non-words**

	*sPvP	*sKvK	IDENT[place]	*sTvT
1 [sTvT]				*
2 [sKvK]		*		
3 [sPvP]	*			

Output of EVAL

Cut-off

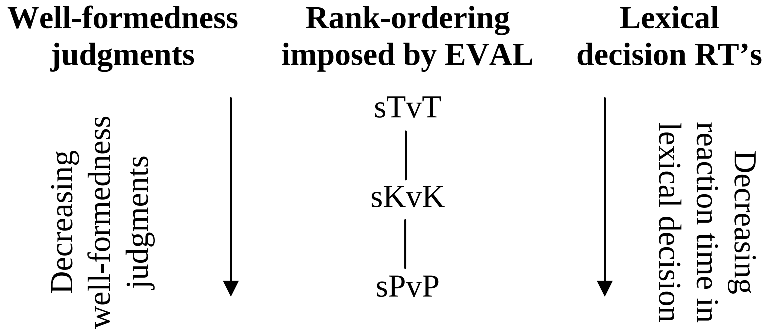
sTvT

sKvK

sPvP

[sTvT]-forms violate the lowest ranked constraint and are therefore rated best by the grammar. [sPvP]-forms violate the highest ranked constraint and are therefore rated worst by the grammar. [sKvK]-forms occupy the intermediate position. From this we get the following predictions about well-formedness judgments: [sTvT]-forms will be rated as most well-formed, then [sKvK]-forms, and finally [sPvP]-forms. About lexical decision reaction times, the following predictions can be made: [sPvP]-forms will be rejected most quickly, [sKvK]-forms more slowly, and [sTvT]-forms will have the slowest rejection times. These predictions are represented graphically in (58).

(58) **Predictions with regard to well-formedness judgments and lexical decision**



In the next section I discuss a series of well-formedness judgment and lexical decision experiments that I conducted to test these predictions. The results of these experiments confirmed that the restriction on [sCvC]-words have an influence on how English speakers process non-words of this form. In particular, the results confirm the predictions shown in (58).

3.2 Word-likeness and lexical decision in English

Inspired by the experiments conducted by Berent *et al*. (see §2.2) for Hebrew, I conducted a similar set of experiments for English. These included both well-formedness judgment experiments and lexical decision experiments. In the experiments, subjects did indeed respond according to the predictions that follow from the analysis developed in §3.1 above. In the well-formedness judgment experiments, subjects rated the possible words ([sTvT]) better than both kinds of impossible words ([sKvK] and [sPvP]). They also distinguished between the two kinds of impossible words, rating [sKvK]-forms better than [sPvP]-forms. In lexical decision they distinguished between possible and impossible words, and between different kinds of impossible words. Non-words of the form [sTvT] were rejected slower than non-words of the form [sKvK] and [sPvP]. But the impossible non-words [sKvK] were also rejected slower than the impossible non-words [sPvP]. These experiments therefore confirmed the predictions of the analysis developed above in §3.1. But the experiments also show that language users make finer well-formedness distinctions than simply between grammatical/possible and ungrammatical/impossible. They also distinguish in the set of impossible words between the more and the less well-formed. Together with the results from Hebrew where we saw that language users distinguish between possible words in terms of well-formedness, these results show that grammar must be able to make multi-level well-formedness distinctions. In the rest of this section I first discuss the well-formedness judgment experiments (§3.2.1), and then the lexical decision experiments (§3.2.2) that I conducted. Before getting into the details of the individual experiments, I will mention those aspects of the experimental design that were shared between the three experiments.

Recordings

The tokens for all experiments were recorded as read by a phonetically trained female speaker of standard American English. Recordings were made directly onto CD in a sound-proofed room. Each token was cut from the speech stream and stored electronically. The intensity of all tokens was equalized.

Token selection

An assumption that underlies all of the experiments that I discuss below is that grammar influences how language users perform well-formedness judgments and lexical decision tasks. This assumption requires that we control for the possible influence of lexical statistics that we know to influence both of these tasks (see the discussion in §1.2 above). The most rigorous control would be to select tokens such that lexical statistics and grammar conflict. Consider the comparison between [sTvT]-forms and [sKvK]-forms. According to the grammatical analysis developed above (§3.1), [sTvT] is less marked than [sKvK]. Grammar favors [sTvT], and we are expecting that [sTvT]-forms should be rated as more well-formed than [sKvK]-forms and that [sTvT]-forms should be detected as non-words more slowly than [sKvK]-forms. If we selected tokens such that the lexical statistics favor [sKvK]-forms over [sTvT]-forms, then lexical statistics would make the opposite predictions – [sKvK] should be rated better and detected more slowly as non-words. If in spite of such a conflict we found that subjects responded according to the predictions of grammar, it could be interpreted as strong evidence that grammar rather than lexical statistics determined their responses.

A less rigorous control for the possible influence of lexical statistics would be to select tokens such that the lexical statistics between tokens types do not differ significantly. Consider again the comparison between [sTvT]-forms and [sKvK]-forms. If the lexical statistics of the [sTvT]-tokens and the [sKvK]-tokens do no differ from each other, then we cannot attribute to lexical statistics any difference in the way in which these tokens types are responded to. If we found that in such a situation subjects do rate [sTvT]-tokens better and detect them more slowly as non-words, we can attribute this to the influence of grammar.

In the selection of tokens for the experiments I first tried to use the more rigorous control. I resorted to the less rigorous control only when this was not possible.

Subjects

The same subjects participated in all three experiments. The subjects were 20 undergraduate students from the University of Massachusetts. All subjects were native speakers of American English, and none of them reported any speech or hearing disabilities. Subjects took part in the experiment for course credit in an introductory Linguistics class. Subjects were tested individually or in groups of up to four.

Order of experiments

There are three experiments and therefore six different orderings possible between the experiments. The order in which the experiments were presented was varied between subjects, so that each of the six possible orderings was presented to roughly an equal number of subjects.

3.2.1 Well-formedness judgment experiments

Following Berent *et al.* I conducted two kinds of well-formedness judgment experiments, namely a gradient well-formedness judgment experiment and a comparative well-formedness judgment experiment. In the gradient well-formedness judgment experiment subjects are presented with individual tokens and rate each token for its word-likeness/well-formedness on a 5-point scale. In the comparative well-formedness judgment experiment subjects are presented with pairs of tokens and select the member of a pair that they deem most well-formed.

3.2.1.1 Gradient well-formedness judgment experiment

In this experiment I presented subjects with a list of tokens that contained non-words of the form [sTvT], [sKvK] and [sPvP]. Subjects were required to rate each token on a 5-point scale where a score of [5] corresponded to a form that they deemed to be very well-formed or very likely to be included in the lexicon of English. A score of [1] corresponded to a form that was not well-formed at all and that was very unlikely to ever be included in the lexicon of English. The three conditions in this experiment and the predictions in each condition are summarized in (59).

The first two conditions represent a comparison between possible words ([sTvT]) and impossible words ([sKvK] and [sPvP]), and the prediction is that the possible words will be rated better than the impossible words. The more interesting comparison is the third one, where the two kinds of impossible words are compared. The prediction there is that, although neither [sKvK] nor [sPvP] is a possible word of English, [sKvK] will be rated better than [sPvP]. Below I first discuss the experimental design (§3.2.1.1.1) and then the results of the experiment (§3.2.1.1.2).

(59) **Gradient well-formedness: Conditions and predictions**

Condition	**Prediction**
T~K	[sTvT] rated better than [sKvK]
T~P	[sTvT] rated better than [sPvP]
K~P	[sKvK] rated better than [sPvP]

3.2.1.1.1 Experimental design

Token selection

In this experiment we are looking for the influence of grammar on well-formedness judgments. It is therefore important to control for other factors that are known to influence well-formedness judgments. In particular, we have to control for the influence of lexical statistics (see discussion in §1.2.1). Non-words that inhabit denser lexical neighborhoods are usually rated as more well-formed than non-words that inhabit sparser lexical neighborhoods. Similarly, non-words with higher phoneme transitional probabilities and/or have higher phoneme transitional probabilities are rated as more well-formed. In the token selection I controlled for the influence of these two kinds of lexical statistics.

Consider first the T~K-condition. In this condition I selected five non-words of the form [sTvT] and five non-words of the form [sKvK]. For each non-word I calculated a lexical neighborhood density[47] and a cumulative bi-phone transitional probability.[48, 49] These tokens were selected such that the mean lexical neighborhood density of the five [sTvT]-tokens did not differ significantly from the mean lexical neighborhood density of the five [sKvK]-tokens. Similarly, the mean cumulative bi-phone probability of the two kinds of tokens did not differ significantly. Now consider the T~P-condition. In this condition I selected five non-word tokens of the form [sTvT] and five non-word tokens of the form [sPvP]. These tokens were also selected such that lexical statistics did not favor any of the token types. Finally, consider the K~P condition. For this condition five non-words of the form [sKvK] and five non-words of the form [sPvP]-were selected. The lexical neighborhood density of the [sKvK]-tokens and [sPvP]-tokens do not differ significantly. However, the [sPvP]-tokens had significantly higher cumulative bi-phone probabilities than the [sKvK]-tokens. In this condition, cumulative bi-phone probabilities therefore favor the [sPvP]-tokens. The table in (60) shows the mean lexical statistics of

---

[47] The lexical statistics were all calculated from the CELEX database (Baayen *et al.,* 1995). This database was “Americanized” before the calculations were done. In the “Americanization” the phonetic transcription of the forms in the database were changed so that they reflect standard American rather than British pronunciation. I am indebted to John Kingston for this.

Lexical neighborhood density 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.

Lexical neighborhood density is calculated as follows: (i) Find all the neighbors for a token. (ii) Sum the log frequencies of all the neighbors. The lexical neighborhood density therefore takes into account both the number of neighbors and their frequencies.

[48] Transitional probabilities were also calculated form the CELEX database (Baayen *et al.*, 1995), following the method used by Vitevitch and Luce (1998, 1999) and Newman *et al*. (1997). Consider the token [sKɔK] as an example. For the sequence [sK] we can calculate the probability of an [s] being followed by [K], and the probability of a [K] being preceded by an [s]. To calculate the probability of [s] being followed by [K]: (i) add up the log frequencies of all tokens that contain an [s]; (ii) add up the log frequencies of all tokens that contain the sequence [sK]; (iii) divide the log frequency of [sK] by the log frequency of [s]. The probability of a [K] being preceded by a [s] can be calculated in a similar manner.

[49] The cumulative bi-phone probability is the product of the individual bi-phone probabilities for a token. This is intended as a measure of the overall probability of the token.

the token types in each of the conditions. (A list of all the tokens and the lexical statistics of each token can be found in the Appendix.)

(60) **Lexical statistics for gradient well-formedness judgment experiment**
(LND = lexical neighborhood density; CBP = cumulative bi-phone probability)

a. T~K-condition

	[sTvT]	[sKvK]	
LND	27.43	23.31	$t(8) = 0.52$, two-tailed $p = 0.62$
CBP	8.86×10^{-9}	6.99×10^{-9}	$t(8) = 0.30$, two-tailed $p = 0.77$

b. T~P-condition

	[sTvT]	[sPvP]	
LND	27.43	22.77	$t(8) = 0.67$, two-tailed $p = 0.52$
CBP	8.86×10^{-9}	3.44×10^{-9}	$t(8) = 1.09$, two-tailed $p = 0.31$

c. K~P-condition

	[sKvK]	[sPvP]	
LND	17.47	22.77	$t(8) = 1.18$, two-tailed $p = 0.27$
CBP	6.46×10^{-10}	3.44×10^{-9}	$t(8) = 3.76$, one-tailed $p = 0.002$[50]

Procedure

The list of tokens that were presented to subjects was constructed as follows: There was a total of 24 test-tokens.[51] Each of these 24 test-tokens was included twice in the list. To

---

[50] This *p*-value is based on a one-tailed *t*-test, assuming that the cumulative bi-phone probability of the [sPvP]-tokens is higher than that of the [sKvK]-token.

this 77 fillers[52] were added so that the final list contained 125 tokens. The list was presented auditorily to the subjects twice so that each test-token was presented four times. There was a break of about five minutes between the two presentations of the list. The list was differently randomized on each presentation. After hearing a token, subjects indicated their rating of the token on an answer sheet by circling a number from [1] to [5]. A score of [1] corresponded to a token that was judged as not very well-formed/very unlikely to ever be included in the lexicon of English. A score of [5], on the other hand, corresponded to a token that was judged to be very well-formed/very likely to be included in the lexicon of English. After a lapse of 5 seconds, the next token was presented. Before the list was presented the first time, 10 filler tokens were presented as practice trials.

3.2.1.1.2 Results

The results were submitted to a repeated measure ANOVA, with mean rating as dependent variables, and markedness (marked~unmarked)[53] and condition (T~K, T~P, K~P) and as independent variables. A main effect of markedness was found both by subjects ($F(1, 19) = 25.01$, $p < 0.000$) and by items ($F(1, 4) = 136.69$, $p < 0.000$). There

---

[51] This number is smaller than the expected number of 30 (3 conditions × 2 token types per condition × 5 tokens per token type), because The reason for this is that the same token is sometimes used in two different conditions.

[52] 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 syllable of a single syllable (Fudge, 1969). They were therefore ill-formed for reasons similar to ill-formedness of the [sKvK] and [sPvP]-tokens.

[53] In each condition, the tokens that are more marked according to the grammar were coded as “marked” and the tokens that are less marked according the grammar were coded as “unmarked”. In the T~K-condition, for instance, [sTvT]-tokens were coded as “unmarked” and [sKvK]-tokens as marked.

was also a significant interaction between markedness and condition both by subject ($F(2, 18) = 19.17$, $p < 0.000$) and by items ($F(2, 3) = 51.69$, $p = 0.005$).

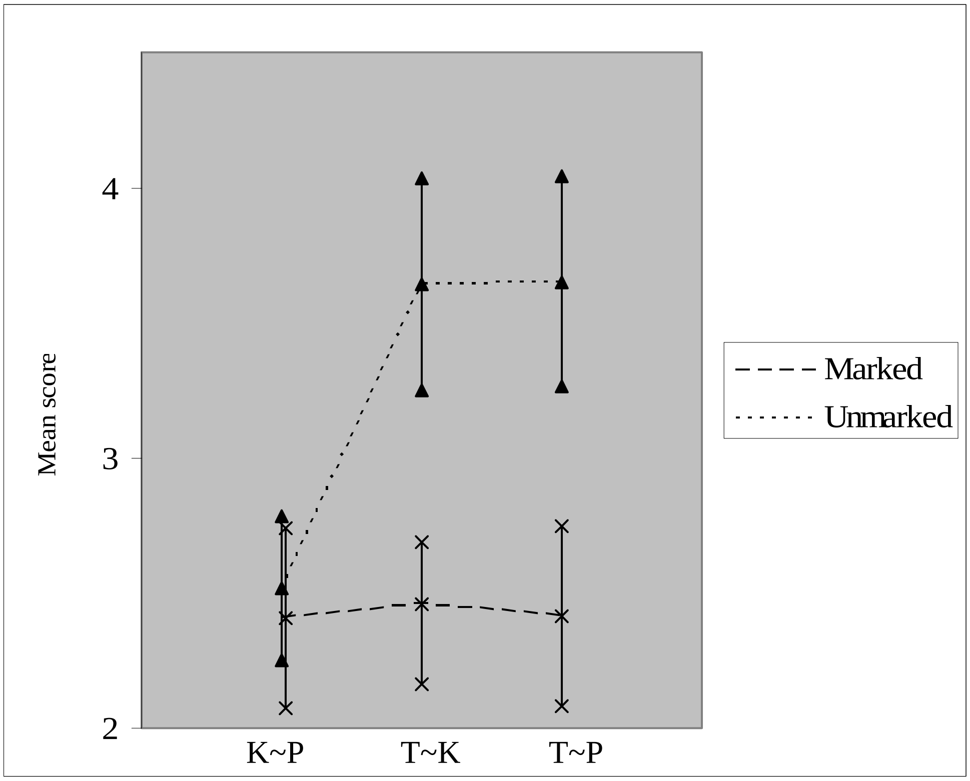
The contrast between the marked and unmarked tokens in each condition was further investigated with one-tailed *t*-tests. (One-tailed tests are called for because I am testing an *a priori* hypothesis about the relationship between well-formedness ratings in each condition.) In the K~P-condition no significant difference was found between the marked [sPvP]-tokens and unmarked [sKvK]-tokens, either by subjects ($t(19) = 0.88$, $p = 0.19$) or by items ($t(8) = 1.29$, $p = 0.12$). However, in the T~K-condition, the unmarked [sTvT]-tokens were rated better than the marked [sKvK]-tokens both by subjects ($t(19) = 5.81$, $p < 0.000$) and by items ($t(8) = 13.61$, $p < 0.000$). Similarly, in the T~P-condition, the unmarked [sTvT]-tokens received higher ratings than the marked [sPvP]-tokens, both by subjects ($t(19) = 5.39$, $p < 0.000$) and by items ($t(8) = 13.40$, $p < 0.000$).

The figures in (61) represent the results graphically, and the results are also summarized in the tables in (62).

Discussion

In this experiment we have very strong evidence that non-words that are less marked according to the grammar are rated as more well-formed than non-words that are more marked. This is confirmed by the highly significant results on the main effect of markedness in the ANOVA’s. More specifically, we also have unequivocal evidence that non-words of the form [sTvT] are rated as more well-formed than non-words of the form [sPvP] and [sKvK]. However, we do not have strong evidence for a preference for [sKvK]-tokens over [sPvP]-tokens. [sKvK]-tokens were rated better than [sPvP]-tokens on average, but the difference in ratings for these two kinds of tokens was small and did not reach significance. Although there is not conclusive evidence that [sKvK] is rated as more well-formed than [sPvP], there is also no evidence to the contrary – i.e. [sPvP]-tokens were not rated better than [sKvK]-tokens. This is particularly relevant since the [sPvP]-tokens were favored over the [sKvK]-tokes in terms of the cumulative bi-phone probabilities (see table (60c) above). Had lexical statistics determined the response pattern, then [sPvP]-tokens should have been rated better. We also have no conclusive evidence that lexical statistics determined the response patterns in the [sKvK]~[sPvP]-condition.

(61) **Gradient well-formedness judgments: Mean scores by subject (with 95%-confidence intervals)**



(62) **a. Overall results (ANOVA's)**

	By subject		**By item**	
Main effect of markedness	$F(1,19) = 25.01$	$p < 0.000$	$F(1,4) = 136.69$	$p < 0.000$
Interaction with condition	$F(2,18) = 19.17$	$p < 0.000$	$F(2,3) = 51.69$	$p = 0.005$

b. Contrasts per condition

	K~P		**T~K**		**T~P**	
	[sKvK]	**[sPvP]**	**[sTvT]**	**[sKvK]**	**[sTvT]**	**[sPvP]**
Mean	2.52	2.41	3.64	2.43	3.65	2.41
By subject[54]	$t(19) = 0.88$, $p = 0.19$		$t(19) = 5.81$, $p < 0.000$		$t(19) = 5.39$, $p < 0.000$	
By item[55]	$t(8) = 1.29$, $p = 0.12$		$t(8) = 13.61$, $p < 0.000$		$t(8) = 13.40$, $p < 0.000$	

3.2.1.2 Comparative well-formedness judgment experiment

In this experiment I presented subjects with non-word token pairs. Each pair had one of the following forms: [sTvT]~[sKvK], [sTvT]~[sPvP], or [sKvK]~[sPvP]. The task of subjects was to select from each pair the member that they considered to be most well-formed/most likely to be included in the lexicon of English. The three conditions in this experiment and the predictions in each condition are summarized in (63).

As in the previous experiment, the first two conditions represent a comparison between possible words ([sTvT]) and impossible words ([sKvK] and [sPvP]), and the prediction is that the possible words will be preferred over the impossible words. The third condition is a comparison between two kinds of impossible words. The prediction there is that, although neither [sKvK] nor [sPvP] is a possible word of English, [sKvK]

---

[54] t-test is paired sample of two means, p-values are one-tailed.

[55] t-test is (non-paired) two sample of means, p-values are one-tailed.

will be preferred over [sPvP]. Below I first discuss the experimental design (§3.2.1.2.1) and then the results of the experiment (§3.2.1.2.2).

(63) **Comparative well-formedness: Conditions and predictions**

Condition	**Prediction**
T~K	[sTvT] preferred over [sKvK]
T~P	[sTvT] preferred over [sPvP]
K~P	[sKvK] preferred over [sPvP]

3.2.1.2.1 Experimental design

Token selection

Consider the T~K-condition as an example. I selected 15 non-word pairs of the form [sTvT]~[sKvK]. The members of a pair were selected so that the cumulative bi-phone probability of the more marked [sKvK]-member was higher than that of the less marked [sTvT]-member. For most of the pairs, the lexical neighborhood density of the more marked [sKvK]-token was also higher than that of the less marked [sTvT]-token.[56] The mean lexical neighborhood density and cumulative bi-phone probability of the more marked [sKvK]-tokens were higher than that of the less marked [sTvT]-tokens. The tokens in the other two conditions were selected in a similar manner. The mean lexical statistics of the tokens in each of the three conditions are given in (64). (The actual tokens used and their lexical statistics are given in the Appendix.)

---

[56] It was not possible to select 15 pairs in which both the lexical neighborhood density and the cumulative bi-phone probability of the more marked [sKvK]-token were higher than that of the less marked [sTvT]-token. Vitevitch and Luce (1998, 1999) have shown that transitional probabilities are more important than lexical neighborhood density in tasks that do not require lexical access. Since well-formedness judgments do not require lexical access, I decided that it was more important to control for transitional probability than for lexical neighborhood density.

This experimental design presents a rigorous test of the hypothesis that grammar in addition to lexical statistics influences well-formedness ratings. In all three conditions, lexical statistics favor the more marked tokens while grammar favors the less marked tokens. For instance, the lexical neighborhood density and cumulative biphone probability of the [sPvP]-tokens were higher than that of the [sTvT]-tokens. Based on this we would expect that [sPvP]-tokens will be rated better. In terms of markedness, the [sTvT]-forms were more well-formed. Based on this we would expect that subjects would rate the [sTvT]-tokens better. Lexical statistics and grammar conflict directly. If the responses of the subjects reflect the influence of grammar rather than lexical statistics, it would count as strong evidence that grammar does influence well-formedness ratings.

(64) **Lexical statistics for comparative well-formedness judgment experiment**
(LND = lexical neighborhood density; CBP = cumulative bi-phone probability)

a. T~K-condition

	[sTvT]	[sKvK]	
LND	16.55	24.21	*t*(14) = 3.32, one-tailed *p* < .003
CBP	1.62×10^{-9}	8.15×10^{-9}	*t*(14) = 3.43, one-tailed *p* < .003

b. T~P-condition

	[sTvT]	[sPvP]	
LND	14.87	21.45	*t*(14) = 4.30, one-tailed *p* < .001
CBP	6.98×10^{-10}	3.17×10^{-9}	*t*(14) = 6.21, one-tailed *p* < .000

((64) continued)

c. K~P-condition

	[sKvK]	[sPvP]	
LND	10.30	21.26	$t(14) = 5.26$, one-tailed $p < .000$
CBP	2.58×10^{-10}	3.18×10^{-9}	$t(14) = 8.88$, one-tailed $p < .000$

Procedure

There were 15 token-pairs in each of the three conditions, resulting in 45 test-pairs. I added 45 filler pairs to this. Only non-words were used in the filler pairs.[57] 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 – i.e. if two tokens occurred in the order [Token 1]~[Token 2] in List 1, then they occurred in the order [Token 2]~[Token 1] in List 2. Both lists were presented auditorily to subjects. 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 instructed that they would hear a pair of non-words, and that their task would be to select the member of each pair that they thought could most likely be

---

[57] The fillers in this experiment were selected using the same criteria as in the previous experiment. See footnote 54.

included in the lexicon of English in the future. 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. Responses were recorded electronically. The next token pair was presented 500 ms after all subjects had responded, or after a time of 5 seconds has elapsed. If a subject did not respond within this time frame, a non-response was recorded.

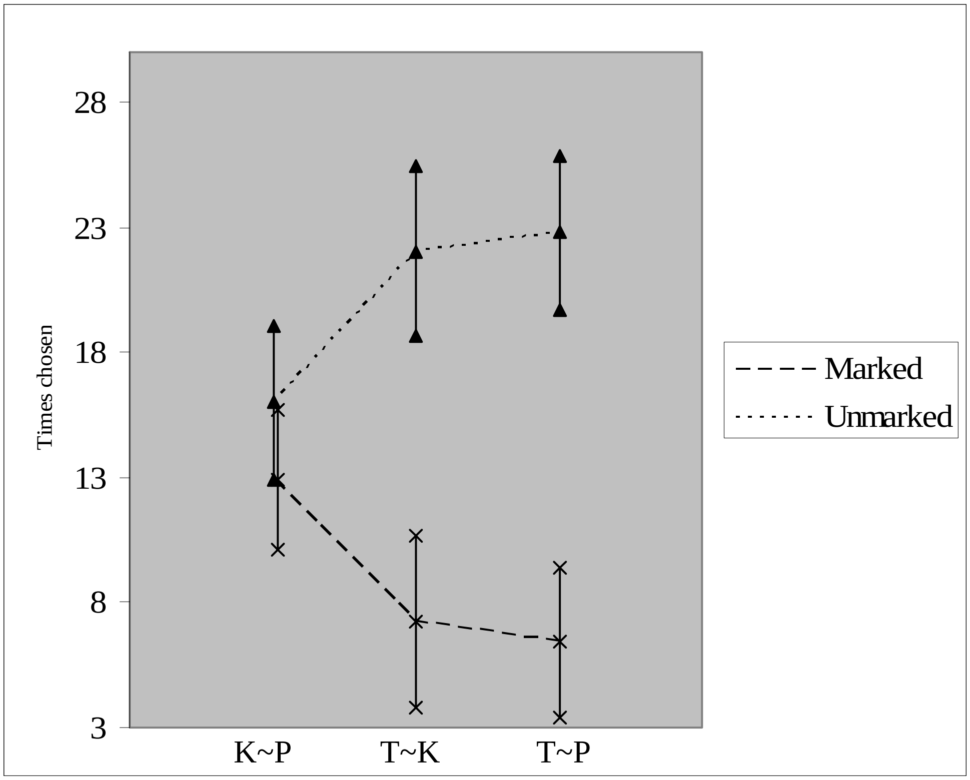
3.2.1.2.2 Results

There were 15 token pairs per condition, each of which was presented twice. There were therefore 30 responses per subject per condition. On every response a subject could either select the more marked or the less marked member. These results 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 both by subjects ($F(1,19) = 23.28$, $p < 0.000$) and by items ($F(1,14) = 188.43$, $p < 0.000$). There was also a significant interaction between markedness and condition both by subjects ($F(2,18) = 10.37$, $p = 0.001$) and by items ($F(2,13) = 23.91$, $p < 0.000$).

The contrast between the marked and unmarked tokens in each condition was further investigated with one-tailed *t*-tests. In the K~P-condition there was an advantage for the less marked [sKvK]-tokens over the more marked [sPvP]-tokens. This difference was significant by items ($t(14) = 1.92$, $p = 0.037$), but not by subjects ($t(19) = 1.12$, $p = 0.14$). In the T~K-condition, the unmarked [sTvT]-tokens were preferred over the marked [sKvK]-tokens both by subjects ($t(19) = 4.54$, $p < 0.000$) and by items ($t(14) = 15.58$, $p < 0.000$). Similarly, in the T~P-condition, the unmarked [sTvT]-tokens were preferred over the marked [sPvP]-tokens, both by subjects ($t(19) = 5.73$, $p < 0.000$) and by items ($t(14) = 13.09$, $p < 0.000$).

The figures in (65) represent the results graphically, and the results are also summarized in the tables in (66).

(65) **Comparative well-formedness judgments: Mean number of choices for the marked and unmarked member in each condition by subject (with the 95%-confidence intervals)**



(66) **a. Overall results (ANOVA's)**

	By subject		By item	
Main effect of markedness	$F(1,19) = 23.28$	$p < 0.000$	$F(1,14) = 188.43$	$p < 0.000$
Interaction with condition	$F(2,18) = 10.37$	$p = 0.001$	$F(2,13) = 23.91$	$p < 0.000$

((66) continued)

b. Contrasts per condition

	K~P		T~K		T~P	
	[sKvK]	**[sPvP]**	**[sTvT]**	**[sKvK]**	**[sTvT]**	**[sPvP]**
By subject[58]	16.0	12.9	22.1	7.3	22.8	6.4
	$t(19) = 1.12$, $p = 0.14$		$t(19) = 4.54$, $p < 0.000$		$t(19) = 5.73$, $p < 0.000$	
By item[59]	21.3	17.2	29.4	9.7	30.4	8.5
	$t(14) = 1.92$, $p = 0.037$		$t(14) = 15.58$, $p < 0.000$		$t(14) = 13.09$, $p < 0.000$	

Discussion

Like the gradient well-formedness experiment, this experiment provides very clear evidence for the effect of markedness in general. This is confirmed by the highly significant results on the main effect of markedness in the ANOVA's. We also have very strong evidence that non-words of the form [sTvT] are preferred over non-words of the form [sKvK] and [sPvP] – this preference was found to be highly significant both by subjects and by items. In this experiment there is also evidence that non-words of the form [sKvK] are preferred over non-words of the form [sPvP]. In the K~P-condition there is an absolute preference for the [sKvK]-tokens both in terms of subjects and in terms of tokens. This preference is significant by items.

In all three of the conditions, the more marked tokens were favored by the lexical statistics. (See (64) above.) Had lexical statistics been responsible for these response patterns, we would have expected a preference for the more marked member. The fact

---

[58] *t*-test is paired sample of two means, *p*-values are one-tailed.

[59] *t*-test is paired sample of two means, *p*-values are one-tailed.

that the opposite preference was found in all three conditions is therefore very strong evidence that grammar influences well-formedness judgments. These results show in particular that grammar takes precedence when grammar and lexical statistics conflict.

Based on the grammatical analysis developed above, I hypothesized the following well-formedness relationship between [sCvC]-forms: |sTvT ™ sKvK ™ sPvP|. In both well-formedness judgment experiments we have clear evidence that non-words of the form [sTvT] are judged to be more well-formed than non-words of the form [sPvP] or [sKvK]. I therefore interpret these two experiments as strongly confirming the sub-hypothesis: |sTvT ™ {sKvK, sPvP}|. For the K~P-condition we found an absolute advantage for the less marked [sKvK]-tokens over the more marked [sPvP]-tokens in both experiments. In the gradient well-formedness experiment this preference did not reach significance. However, in the comparative well-formedness judgment experiment, it was found to be significant by items. This gives evidence for the sub-hypothesis |sKvK ™ sPvP|. In general I interpret the results of these experiments as confirmation of the hypothesized well-formedness relation between [sCvC]-forms. The lexical decision experiment discussed below provides additional confirmation of this hypothesis.

3.2.2 Lexical decision experiment

In this experiment I presented subjects auditorily with a list of words and non-words. The task of subjects was to discriminate between words and non-words. The basic 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 quicker the token will be rejected. The non-words that the listeners were presented with included tokens of the form [sTvT],

[sKvK] and [sPvP]. The three experimental conditions and the predictions for the conditions are summarized in (67) below.

(67) **Lexical decision: Conditions and predictions**

Condition	**Prediction**
T~K	RT([sTvT]) > RT([sKvK])
T~P	RT([sTvT]) > RT([sPvP])
K~P	RT([sKvK]) > RT([sPvP])

As in the previous experiments, the first two conditions represent a comparison between possible ([sTvT]) and impossible words ([sKvK] and [sPvP]), and the prediction is that the possible words will be rejected more slowly than impossible words. The third condition is a comparison between two kinds of impossible words. The prediction is that, although neither [sKvK]- nor [sPvP]-forms are possible words of English, [sKvK]-forms will be rejected more slowly than [sPvP]-forms. This was confirmed by the results of the lexical decision experiment. In the rest of this section I will first discuss the design (§3.2.2.1), and then the results (§3.2.2.2) of the lexical decision experiment.

3.2.2.1 Experimental design

Token selection

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]. The tokens were selected such that the mean lexical neighborhood density and the mean cumulative bi-phone probability of the [sTvT]-tokens and the [sKvK]-tokens did not differ significantly. Lexical 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 mean lexical statistics of the tokens in each of the three conditions are given in (68). (The actual tokens used and their lexical statistics are given in the Appendix.)

(68) **Lexical statistics for comparative well-formedness judgment experiment**
(LND = lexical neighborhood density; CBP = cumulative bi-phone probability)

a. T~K-condition

	[sTvT]	[sKvK]	
LND	31.58	27.24	*t*(8) = 0.78, two-tailed *p* > .45.
CBP	4.85×10^{-8}	6.18×10^{-9}	*t*(8) = 1.81, two-tailed *p* > .11

b. T~P-condition

	[sTvT]	[sPvP]	
LND	31.58	26.56	*t*(8) = 1.00, two-tailed *p* > .34
CBP	4.85×10^{-8}	2.54×10^{-9}	*t*(8) = 1.99, two-tailed *p* > .08

c. K~P-condition

	[sKvK]	[sPvP]	
LND	14.95	14.26	*t*(8) = 0.10, two-tailed *p* > .92
CBP	8.08×10^{-10}	8.50×10^{-10}	*t*(8) = 0.06, two-tailed *p* > .95

Procedure

The list of tokens that were presented to subjects was constructed as follows: There were a total of 27 test-tokens.[60] Each of these 27 test-tokens was included once in the list. To

---

[60] 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.

this 76 fillers[61] were added so that the final list contained 103 tokens. These tokens were presented auditorily to subjects. 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 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 the 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.

3.2.2.2 Results

The response times were recorded starting at the onset of a stimulus. Before analyzing the data I subtracted the duration of every stimulus from the recorded response time. The resulting measure represents how long after (or before) the end of the stimulus a subject recorded a response. In the rest of the discussion I will refer to this measure (recorded response time minus token duration) as "response time". The response times for each subject were normalized, 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

---

[61] The fillers in this experiment were selected using the same criteria as in the previous experiments. See footnote 54.

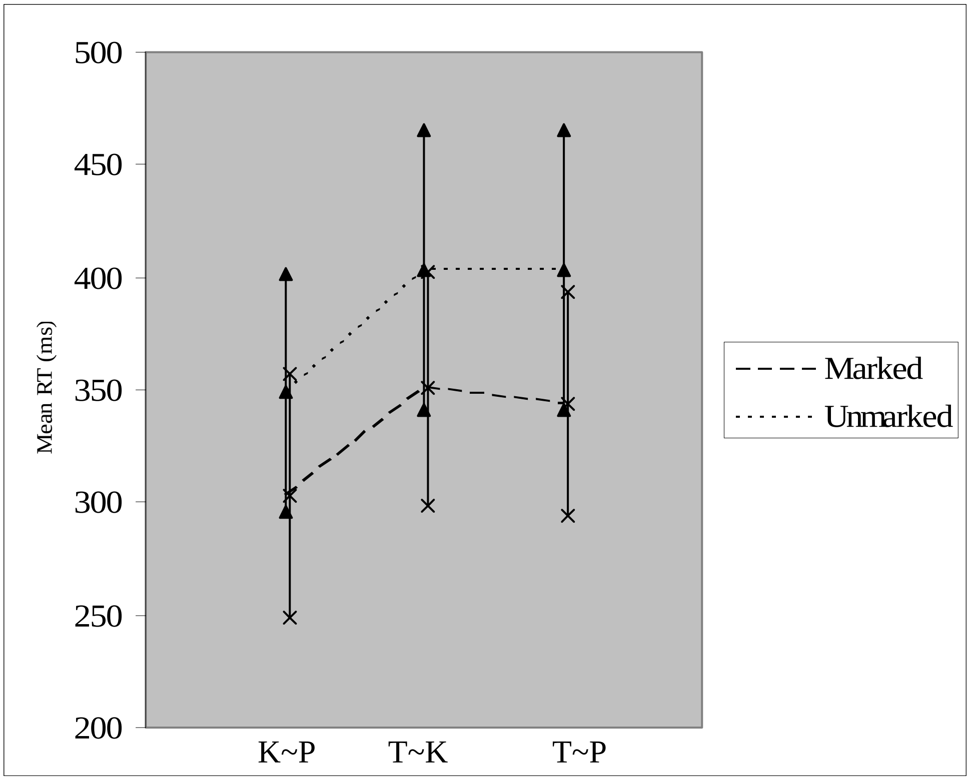
responses were included in the analysis. Exclusion of outliers and incorrect responses resulted in exclusion of only 8% of the total responses.

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 ($F(2, 13) = 0.08$, $p = 0.92$).

The contrast between the marked and unmarked tokens in each condition was further investigated with one-tailed *t*-tests. In the K~P condition the more marked [sPvP]-tokens had shorter reaction times than the less marked [sKvK]-tokens. This difference was significant both by subjects ($t(19) = 3. 79$, $p < 0.000$) and by items ($t(8) = 2.15$, $p = 0.03$). In the T~K-condition the more marked [sKvK]-tokens had shorter reaction times than the less marked [sTvT] tokens. This difference was significant by subjects ($t(19) = 3.40$, $p < 0.002$) but not by items ($t(8) = 1.63$, $p = 0.07$). In the T~P-condition the more marked [sPvP]-tokens were also rejected more quickly than the less marked [sTvT]-tokens. This difference was significant by subjects ($t(19) = 4. 20$, $p < 0.001$) but not by items ($t(8) = 1.55$, $p = 0.08$).

The figures in (69) represent the results graphically, and the results are also summarized in the tables in (70).

(69) **Lexical decision: Mean RT's in ms by subject (with 95% confidence intervals)**



(70) **a. Overall results (ANOVA's)**

	By subject		By item	
Main effect of markedness	*F*(1,19) = 21.68	*p* = 0.001	*F*(1,4) = 2.584	*p* = 0.18
Interaction with condition	*F*(2,18) = 0.58	*p* = 0.572	*F*(2,3) = 0.08	*p* = 0.92

b. Contrasts per condition

	K~P		T~K		T~P	
	[sKvK]	**[sPvP]**	**[sTvT]**	**[sKvK]**	**[sTvT]**	**[sPvP]**
By subject[62]	348.98	303.33	403.53	350.45	403.53	344.01
	t(19) = 3.78,	*p* < 0.001	*t*(19) = 3.39,	*p* < 0.002	*t*(19) = 4.20,	*p* < 0.001
By item[63]	345.27	303.86	404.52	351.55	404.52	348.18
	t(8) = 2.15,	*p* = 0.03	*t*(8) = 1.62,	*p* = 0.07	*t*(8) = 1.55,	*p* = 0.080

---

[62] *t*-test is paired sample of two means, *p*-values are one-tailed.

[63] *t*-test is (non-paired) two sample of means, *p*-values are one-tailed..

Discussion

This experiment provides clear evidence that non-words that are more well-formed according to grammar are rejected more slowly than non-words that are less well-formed. This is confirmed by the significant main effect in the subjects ANOVA. It is also confirmed by the *t*-tests for each of the three individual conditions. In this experiment the lexical statistics did not differ significantly between the token types in each condition, so that it is not expected that the lexical statistics should significantly influence the reaction times. I therefore interpret the differences in reaction times that were found as evidence that grammar influences reaction times. Based on the grammatical analysis that I developed in §3.1 above, I hypothesized that the reaction time for the three kinds of non-words would be related as follows: RT([sTvT]) > RT([sKvK]) > RT([sPvP]). This hypothesis was confirmed by the results of the lexical decision experiment. Under the assumption that a non-word is rejected more quickly the less well-formed it is, this could be interpreted as evidence for the following well-formedness relation between the three kinds of [sCvC]-tokens: |sTvT ™ sKvK ™ sPvP|. The results of this experiment serve as strong evidence in favor of the grammatical analysis of these forms developed in §3.1 above.

4. Considering alternatives

In this section I consider alternative accounts for the response patterns observed in the well-formedness judgment and lexical decision experiments discussed above. Two kinds of alternatives are considered. First, I discuss the claim that the response patterns result from the influence of lexical statistics rather than from grammar. Secondly, I consider several alternative grammatical accounts for the response patterns. I will show in particular that a rank-ordering model of EVAL is more suited to account for the response patterns than a classic OT grammar (Prince and Smolensky, 1993), a stochastic OT grammar (Boersma, 1998, Boersma and Hayes, 2001), or a grammar with variable constraint rankings (Anttila, 1997, Anttila and Cho, 1998, Reynolds, 1994, Reynolds and Sheffer, 1994).

4.1 Grammar or lexical statistics?

One of the central assumptions behind the interpretation of the experimental results in §2 and §3 above is that the response patterns observed in these experiments stem from the influence of grammar on phonological processing. However, it is firmly established that lexical statistics influence phonological processing (see the discussion in §1.2 above). Therefore, it is necessary to consider an alternative account for the response patterns – an account that attributes these patterns not to grammar but to lexical statistics. Below I first discuss the Hebrew experiments of Berent *et al*. (§4.1.1) and then the English experiments (§4.1.2), showing that lexical statistics had little or no effect on the response patterns in these experiments.

4.1.1 Hebrew and the OCP

Because no word usage frequency counts are available for Hebrew, Berent *et al*. could not calculate lexical neighborhood density or transitional probabilities for the tokens used in their experiments. It is therefore not possible to show definitively that these two types of lexical statistics did not influence their results. Even so, there is some evidence that is very suggestive of the fact that their results cannot be attributed to lexical statistics.

Consider first the well-formedness judgment experiments of Berent *et al*. The data discussed above in §2.2.1 come from a study by Berent and Shimron (1997). They did not report on the lexical statistics of their tokens. However, Berent *et al*. (2001a) conducted the exact same kind of well-formedness judgment experiments with the same experimental design as Berent and Shimron (1997), and attained results that were exactly comparable to those of Berent and Shimron (1997) discussed above. Berent *et al*. (2001a) did report on some lexical statistics of their tokens. I will therefore discuss the results of Berent *et al*. (2001a) here, and make the assumption that what is true about their results can be transferred to the results of Berent and Shimron (1997).

A reminder of the design and results of the experiments: Tokens consisted of forms that did not correspond to actual Hebrew words. The tokens were of three kinds – forms with no identical stem consonants ([QiSeM], no-geminate forms), forms with identical consonants in the last two stem positions ([QiSeS], final-geminate forms), and forms with identical consonants in the initial two stem positions ([QiQeS], initial-geminate forms). Based on the OT analysis developed for these forms, it was predicted that they are related as follows in terms of their well-formedness: |No-gemination ™ Final-gemination ™ Initial-gemination|. This is indeed how these forms were rated by the subjects in the experiments of Berent *et al*. What we need to show is that this rating does not correspond to the lexical statistics of these forms – i.e. that the lexical statistics did not favor no-geminate forms over final-geminate forms, and final-geminate forms over initial-geminate forms. This cannot be shown for the comparison between initial-geminate forms and the other two kinds of forms. However, it can be shown for the comparison between no-geminate forms and final-geminate forms.

Berent *et al*. (2001a) did not calculate any lexical statistic that can be correlated to lexical neighborhood density. But they did calculate a statistic that corresponds closely to transitional probability. They compiled a list all of the productive stems in the Even-Shoshan Hebrew dictionary (1993). This resulted in a list of 1,412 forms. Geminates were treated as tri-consonantal in this list – the fact that the root /S-M/ is productive in Hebrew, led to the inclusion of the form |S-M-M| in this list. They then used this list to calculate the positional type bigram frequency of their experimental forms by adding the frequency of their initial C_1C_2 bigram, their final C_2C_3 bigram, and the bigram of the initial and final consonants, C_1-C_3 (Berent *et al*., 2001a:28). In this calculation geminates were therefore treated as tri-consonantal forms. In this way a cumulative bigram frequency index can be calculated for each of the tri-consonantal forms used in their experiment. This is comparable to non-frequency weighted transitional probabilities counts (see §1.2.1).[64] The results of their calculation are shown in the table in (71).

(71) **Summed type positional bigram frequency of the forms used in the well-formedness judgment experiments by Berent *et al*. (2001a)**

Type	Example	Average summed bigram frequency
Initial-gemination	Q-Q-S	6.041
Final-gemination	Q-S-S	11.58
No-gemination	Q-S-M	9.33

---

[64] It would, of course, have been better if these counts could have been weighted according to the usage frequency of the items in their list of stems. However, such counts are not available for Hebrew. Also, recall the result of Bailey and Hahn (Bailey and Hahn, 1998) for transitional probabilities in English discussed in §1.2.1 – they compared frequency weighted transitional probabilities with non-frequency weighted transitional probabilities and found a very high correlation (r^2 = .96).

Berent *et al*. report that the differences in summed bigram frequencies between the three types of tokens were significant. Not surprisingly, the bigram frequency of initial-geminates was the lowest – this is expected since there are very few initial-geminates in the lexicon of Hebrew. It is therefore possible that the rejection of initial-geminates could be ascribed to the fact that these forms all had lower average summed bigram frequencies. However, the average summed bigram frequency of the final-geminate forms was higher than that of the no-geminate forms. If the bigram frequencies determined the ratings assigned to these forms, then we would have expected final-geminate forms to be rated better than no-geminate forms. The fact that the opposite was found serves as strong evidence that the ratings cannot be attributed to the effect of bigram frequencies, and by extension therefore probably not to other kinds of lexical statistics.

Now consider the lexical decision experiments of Berent *et al*. (2001b) discussed in §2.2 above. In this experiment they presented subjects with the same kinds of tokens as that used in the well-formedness judgment experiments. Based on the OT analysis developed for these forms, the prediction was that the rejection times for the different token types would be related as follows: RT(No-geminate) > RT(Final-geminate) > RT(Initial-geminate). The results of the experiment of Berent *et al*. (2001b) did not correspond exactly to this prediction.[65] For instance, they found contra the predictions that final-geminate forms were responded to slower than no-geminate forms. Even though this goes against the predictions of the analysis, we can still consider whether the response pattern can be attributed to the influence of lexical statistics. Using the same

---

[65] On their results and on why it deviated from the predictions, see the discussion in §2.2.2 above.

method described just above, Berent *et al.* calculated summed bigram frequencies for the no-geminate and the final-geminate forms used in their experiments. These statistics are reported in the table in (72). They found that the bigram frequency for these two kinds of tokens did not differ significantly. Therefore, the difference in response time associated with these two kinds of tokens cannot be attributed to their bigram frequencies.

(72) **Summed type positional bigram frequency of the forms used in the lexical decision experiments by Berent *et al*. (2001b)**[66]

Type	**Example**	**Average summed bigram frequency**
Final-gemination	Q-S-S	10.63
No-gemination	Q-S-M	10.93

Based on this discussion we can conclude that it is very unlikely that response patterns in the experiments of Berent *et al.* are the result of the influence of lexical statistics.

4.1.2 [sCvC]-forms in English

In the design of the experiments that I conducted on the processing of [sCvC]-forms in English I did control for the potential influence of both lexical neighborhood density and transitional probabilities on the results of the experiments. In all of the experiments I selected tokens such that these lexical statistics either did not differ between the different token kinds, or such that the lexical statistics predicted response patterns opposite to that predicted by the grammatical analysis that I developed. The fact that the results of the experiments confirmed the predictions of the grammatical analysis therefore counts as

---

[66] Berent *et al.* (2001b) unfortunately do not report the bigram frequency for the initial-geminate tokens that they used in their experiment.

strong evidence in favor of the claim that these results do reflect the influence of grammar rather than of lexical statistics.[67]

Consider the well-formedness judgment experiments. In these experiments subjects had to rate non-words of the form [sTvT], [sKvK] and [sPvP]. Based on the grammatical analysis that I developed for these forms (§3.1), the prediction was that these forms would be rated as follows: |sTvT ™ sKvK ™ sPvP|. The lexical statistics either favored the less well-formed tokens (for instance, favoring [sPvP] over [sKvK]), or did not differ significantly between the token types. Lexical statistics therefore either made predictions that conflicted with the grammatical analysis, or predicted no difference between token types. In spite of these lexical statistics, the experimental results show that subjects did indeed rate the forms as predicted by the grammatical analysis.

I also conducted a lexical decision experiment. The predictions here was that the reaction times to the different token types would be related as follows: RT([sTvT]) > RT([sKvK]) > RT([sPvP]). Again, lexical statistics either predicted no difference between the token types or made predictions counter to these. In spite of the lexical statistics, it was found that subjects did respond to the tokens as predicted based on the grammatical analysis.

This alone is enough to show that the results of these experiments cannot be attributed to the influence of lexical statistics. However, it is possible to show this even more conclusively. We can perform regression analyses on the response data in each experiment, using the lexical statistics as the independent variable. In this manner it is

---

[67] The details about the tokens used and the lexical statistics are discussed in the experimental design sections above, and I will not repeat these here (see §3.2 above).

possible to quantify the amount of variation in the response data that is accounted for by the lexical statistics. The table in (73) shows the results of these analyses.

(73) **Regression on response data from the three experiments on [sCvC]-tokens**

	Lexical Neighborhood Density	**Cumulative Bi-phone Probability**
Gradient well-formedness	$r^2 = 0.04$ $F(1,43) = 2.02$, $p > 0.05$	$r^2 = 0.02$ $F(1,43) = 0.86$, $p > 0.05$
Comparative well-formedness	$r^2 = 0.04$ $F(1,43) = 2.02$, $p > 0.05$	$r^2 = 0.02$ $F(1,43) = 0.86$, $p > 0.05$
Lexical decision	$r^2 = 0.18$ $F(1,20) = 4.28$, $p > 0.05$	$r^2 = 0.10$ $F(1,20) = 2.33$, $p > 0.05$

Lexical statistics therefore do not contribute significantly to the variance observed in any of the three experiments. Since the results of the experiments cannot be attributed to lexical statistics, I interpret the results as strong evidence that grammar does influence the processing of non-words.

4.2 Alternative grammatical analyses

If the results of the experiments do reflect the influence of grammar, then our theory of grammar should be able to account for these data. In the discussion above I claim that these data can be accounted for within the rank-ordering model of EVAL. In this section I will consider alternative grammatical accounts, and show that the rank-ordering model of EVAL accounts better for these data. I will first consider a classic OT grammar (Prince and Smolensky, 1993), showing it cannot account for these data (§4.2.1). This serves as motivation for the extensions to a classic OT grammar that I propose in this dissertation. After that I will consider stochastic OT grammars (Boersma, 1998, Boersma and Hayes, 2001), and OT grammars with variable constraint rankings (Anttila, 1997, Anttila and

Cho, 1998, Reynolds, 1994, Reynolds and Sheffer, 1994). These OT grammars were designed specifically to handle non-categorical data and therefore seem particularly well suited to the kind of data discussed in this chapter. However, I will show that these models face certain practical and conceptual problems that the rank-ordering model of EVAL can more easily overcome.

4.2.1 Classic OT

I argue for two extensions to the architecture of a classic OT grammar in this dissertation (Chapter 1 §1). First, I claim that EVAL can compare candidates that are not related to each other via a shared input. Secondly, I claim that EVAL imposes a harmonic ordering on the full candidate set, rather than just to distinguish the best candidate from the rest of the candidates. In this section I will show that both of these extensions are necessary to account for the results of the experiments discussed in this chapter.

Consider first the claim that EVAL can compare morphologically unrelated candidates. A classic OT grammar is a function from an input to an output so that EVAL can compare only candidate output forms that are generated by GEN for the same input. There is no direct way in which to relate morphologically unrelated forms to each other in terms of the grammar. This is of course true not only of classic OT, but of all grammars in the classic generative tradition. Since generative grammars are designed as functions that map a single input onto a single output, they cannot easily relate outputs that are derived from unrelated inputs to each other.

In grammars that allow no such inter-output comparison it is in principle not possible to account for the experimental results discussed above. All of these experiments imply comparison between different non-words that are not morphologically related to each other. If we want to account for the response patterns in terms of grammar we need a grammar that is capable of comparing morphologically unrelated forms. In terms of an OT grammar, we need EVAL to be able to compare morphologically unrelated forms.

Now that the need for the first extension to an OT grammar has been established, what about the second extension? Is it also necessary to allow EVAL to impose a rank-ordering on the full candidate set? Let us consider what would happen if we allowed EVAL to compare morphologically unrelated forms but not to rank-order the full candidate set. EVAL can now compare a set of morphologically unrelated forms to each other (like the three kinds of tokens in the Hebrew experiments and the three kinds of tokens in the English experiments). However, EVAL will make only one distinction in these comparison sets, namely between the best form and the non-best forms. Even though EVAL can compare the three kinds of tokens used in the Hebrew experiments or the three kinds of tokens used in the English experiments, it is incapable of imposing a three-level harmonic ordering on them. This contrasts with the results of these experiments, which showed that language users do make a three level well-formedness distinction.

Since language users do make multi-level well-formedness distinctions we need a theory of grammar that can also do this. A theory that distinguishes only between the best and the rest cannot adequately account for the response patterns observed in the experiments.

In order to account for the response patterns observed in the experiments, both extensions to the classic OT model are necessary. Since language users can compare forms that are not related to each other via a shared input, EVAL should also be able to do this. Since language users make finer distinctions than simply between the best and the non-best candidates, EVAL should also be able to make such multi-level well-formedness distinctions.

4.2.2 Stochastic OT

Stochastic OT grammars (Boersma, 1998, Boersma and Hayes, 2001) were developed to account for variable phenomena. The response data discussed in this chapter are variable. It would therefore seem that stochastic OT should be able to account for these variable response data. However, in this section I will show that these models face both conceptual and principled problems with regard to the data discussed in this chapter. I will point out three kinds of problems. The first two are of a conceptual nature, and I refer to these problems as the "which-values" and the "grammar-alone" problems (see also Chapter 5 §3.1.1). The third problem is a more principled problem, and focuses and the problems faced by these grammars with relative well-formedness difference between possible and impossible words.

The which-values problem. The rank-ordering model of EVAL that I propose in this dissertation generates information only about relative well-formedness relations between forms. Consider the English example discussed in §3. From the rank-ordering model we get the information, for instance, that [sTvT]-forms are more well-formed than [sPvP]-forms. However, no information is generated about the absolute size of the well-formedness difference between these two kinds of forms. From this follows the prediction that [sTvT]-forms will be rated better than [sPvP]-forms. But no prediction is made about how much better [sTvT]-forms will be rated. Similarly, the prediction is that [sTvT] non-words will be rejected more slowly than [sPvP] non-words. But no prediction is made about how much the rejection times between these two kinds of tokens will differ. This represents an important difference between the rank-ordering model of EVAL and stochastic grammars. A stochastic OT grammar makes predictions about the absolute size of well-formedness differences between tokens (Hayes, 1997, 1998). I will first explain how a stochastic grammar can make predictions about absolute well-formedness differences, and then return to the problems that follow from this ability of these grammars.

In a stochastic OT grammar, constraints are ranked along a continuous ranking scale. Every constraint has a basic ranking value along this scale. The actual point where a constraint is ranked along the continuous ranking scale is not equivalent to its basic ranking value. A stochastic OT grammar includes a noise component – on every evaluation occasion a (positive or negative) random value is added to the basic ranking value of every constraint. The result of this addition determines the precise place where that constraint will be ranked along the continuous scale on the particular evaluation occasion.[68] As an example, consider the constraints *sTvT and *sPvP from the discussion in §3 above. Since [sPvP]-forms are more marked than [sTvT]-forms, the basic ranking value of *sPvP will be higher than that of *sTvT. For argument's sake, let us assume that the basic ranking value of *sPvP is 100, and the basic ranking value of *sTvT is 98. On any given evaluation occasion the actual ranking value of *sPvP will be selected from values distributed normally around 100, and the actual ranking value of *sTvT from values distributed normally around 98. Since the mean of the actual ranking values of

---

[68] The random value added to the basic ranking value has a normal distribution, with zero as its mean, and some arbitrarily chosen standard deviation that is set at the same value for all constraints. See Boersma and Hayes (2001).

*sPvP is higher than the mean of the actual ranking values of *sTvT, it is more likely that the actual ranking will be ||*sPvP o *sTvT|| than the other way around. But the opposite ranking is also possible – for instance, when a large enough negative noise value is added to the basic ranking value of *sPvP and a large enough positive value to the basic ranking value of *sTvT. By changing the distance between the basic ranking values of these two constraints, the likelihood of the two rankings ||*sPvP o *sTvT|| and ||*sTvT o *sPvP|| can be controlled to the finest detail.

Let us consider how such a stochastic model of OT can account for well-formedness judgments. Suppose that [sTvT]-forms are rated on average twice as good as [sPvP]-forms. The basic ranking values of *sTvT and *sPvP can be set such that the ranking ||*sPvP o *sTvT|| is twice as likely as the ranking ||*sTvT o *sPvP||. In 67% of all comparisons between an [sTvT]-form and an [sPvP]-form, these forms will be rated as follows |sTvT ™ sPvP|. In the other 33% of comparisons the opposite harmonic ordering will be imposed on the forms – i.e. |sPvP ™ sTvT|. The actual average well-formedness ratings of these forms will then correspond to the likelihood of the different rankings between the constraints.

Unlike the rank-ordering model of EVAL that I am proposing, a stochastic OT grammar can do more than make predictions about relative well-formedness judgments. It can make predictions about the absolute size of the difference in well-formedness ratings between different forms. And since the basic ranking values of constraints are distributed along a continuous scale, it is possible to model any difference in ratings by varying the distance between the basic ranking values of two constraints. In a similar way the differences in reaction times in lexical decision experiments can be modeled very accurately in a stochastic OT grammar.

It seems that a stochastic OT grammar can account better for the well-formedness judgment and lexical decision data. However, this is only an apparent advantage of these grammars. In actuality, the fact that such a grammar makes predictions about the absolute size of the well-formedness difference between forms presents it with a problem: Which absolute values must be modeled? Consider again the data on the English [sCvC]-experiments. The tables in (74) contain a selection of the data from the well-formedness judgment experiments discussed in §3.2.

(74) **Comparison between [sTvT]-forms and [sPvP]-forms**

a. **Gradient well-formedness judgment experiment** (see §3.2.1.1)

	[sTvT]	[sPvP]
Mean score	3.65	2.41
Score ratio[69]	.60	.40

b. **Comparative well-formedness judgment experiment** (see §3.2.1.2)

	[sTvT]	[sPvP]
Mean number of preferences	22.8	6.4
Preference ratio[70]	.78	.22

The advantage of the less marked [sTvT]-forms over the more marked [sPvP]-forms is quite different in these two experiments. It is possible to model the results of the

---

[69] This ratio is calculated as follows: Mean score of *x* / (Mean score of *x* + Mean score of *y*).

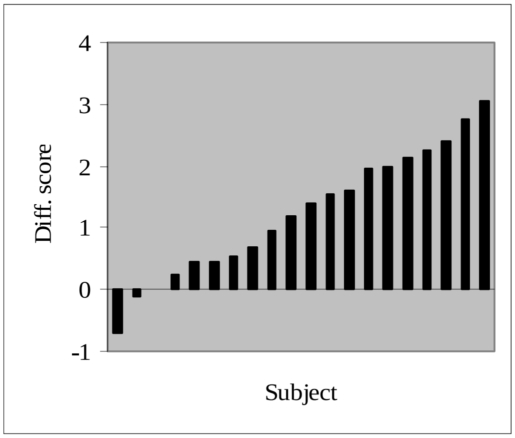
[70] This ratio is calculated as follows: Number of *x*-preferences / (Number of *x*-preferences + Number of *y*-preferences).

gradient well-formedness judgment experiment by placing *sKvK and *sTvT on the continuous ranking scale such that the ranking ||*sPvP o *sTvT|| will be selected 60% of the time and the ranking ||*sTvT o *sPvP|| 40% of the time. But this will then make the wrong predictions about the results of the comparative well-formedness judgment experiment. For this experiment we need the ranking ||*sPvP o *sTvT|| 79% of the time, and the ranking ||*sTvT o *sPvP|| only 21% of the time. It is not possible to model the performance in these two experiments with the same grammar. An account within a stochastic model will have to assume that the same subjects use different grammars in the different experiments.

This same problem of the stochastic models becomes evident also when we compare the differences between subjects. I will discuss the gradient well-formedness judgment experiment for English here as an example, but the same point can be made with the comparative well-formedness judgment experiment or the lexical decision experiment. In (75) I represent the difference score for the T~P-condition for each of the 20 subjects that took part in the gradient well-formedness experiment. This difference score is calculated as follows: For each subject I subtracted his/her mean rating for the [sPvP]-tokens from his/her mean rating for the [sTvT]-tokens.

The figure in (75) shows clearly that subjects differ widely in how much they rate [sTvT]-forms better than [sPvP]-forms. The ratio of a 60%-preference for [sTvT]-forms showed in the table in (74a) above is only the average across subjects. If we model our grammar based on this 60%-preference, we are therefore modeling a very abstract entity – some assumed average grammar across all of the subjects who took part in the experiment. However, it is debatable how real such an “average grammar” is. As an alternative, we could calculate the [sTvT]-preference for every subject individually and then model the grammar for every individual subject based on his/her individual [sTvT]-preference. But this tips the scale too far in the other direction. Now the idea of some shared grammar among the subjects (as members of the same linguistic community) is lost.

(75) **Difference scores by subject for the T~P condition in the gradient well-formedness judgment experiment**



The grammar-only problem. The second conceptual problem faced by the stochastic models is closely related to the first. In the stochastic models all aspects of the variation in the performance of the subjects are modeled directly in the grammar. Implicit to this is the assumption that grammar alone is responsible for the performance of the subjects. We know that this is not the case. At least some of the variation in the ratings results not from grammar but from other factors. The difference in the results of the two well-formedness judgment experiments shows that the specific task, for instance, also contributes to the variation. And we know that factors such as lexical statistics also contribute (see discussion in §1.2.1 above). The problem with a stochastic grammar is that it tries to account for all aspects of the variation in the data through the grammar – it does not allow enough room for the influence of other factors.

In the rank-ordering model of EVAL these problems are avoided. The rank-ordering model of EVAL predicts that grammar will rate [sTvT]-forms as better than [sPvP]-forms. However, it does not dictate the size of the [sTvT]-preferences. The actual size of the difference is the result of a complex interaction with many other variables, including task specific properties and differences between individuals. Within the rank-ordering model of EVAL the same grammar can be assumed, both for the gradient well-formedness judgment experiments and the comparative well-formedness judgment experiment. The difference in the results between these experiments can then be ascribed to factors other than grammar.

Relative well-formedness differences between possible and impossible words. [sTvT]-forms are possible words in English while [sPvP]-forms are not. In spite of this we saw that [sTvT]-forms were not rated infinitely better than [sPvP]-forms. For instance, in the comparative well-formedness judgment experiment [sPvP] was chosen over [sTvT] 22% of the time (see table (74b) above). Let us simplify by considering only the constraints *sTvT, *sPvP and IDENT[place]. The observed well-formedness difference between [sTvT]-forms and [sPvP]-forms can be captured in a stochastic grammar by assuming that we have the ranking ||*sPvP o *sTvT|| 78% of the time, and the ranking ||*sTvT o *sPvP|| 22% of the time. But this causes problems with the distinction between possible and impossible words.

English does not allow words of the form [sPvP]. This implies the following ranking between *sPvP and IDENT[place]: ||*sPvP o IDENT[place]|| (an /sPvP/-input will then be mapped unfaithfully onto an output such as [sPvT]). There is no variation in this regard in English – an [sPvP]-form is never a possible word. The absence of variation means that the basic ranking values of *sPvP and IDENT[place] should be different enough that the ranking ||*sPvP o IDENT[place]|| will (practically) never be inverted by the addition of random noise to these basic ranking values.

But now are we predicting that variation should be observed in the way that English treats an /sTvT/-input. [sPvP] is preferred over [sTvT] 22% of the time, meaning that the ranking ||*sTvT o *sPvP|| is observed 22% of the time. /sPvP/ is never mapped faithfully, meaning that the ranking ||*sPvP o IDENT[place]|| is observed (practically) 100% of the time. This implies that the ranking ||*sTvT o IDENT[place]|| should be observed (at least) 22% of the time. And whenever this ranking is observed, then an /sTvT/-input would not be allowed to map faithfully onto itself. At least 22% of the time a word such as *state* should therefore be pronounced not as [steɪt] but as something like [steɪk] or [steɪp].[71] The rank-ordering model of EVAL avoids this problem. In this model the ranking between the constraints is always the same, namely ||*sPvP o IDENT[place] o *sTvT o Cut-off||. Because the cut-off occurs at the bottom no variation will be observed. /sTvT/ will always map faithfully onto itself and /sPvP/ will never map faithfully onto itself.

---

[71] I have sketched the argument here assuming that the ranking ||*sPvP o IDENT[place]|| is fixed. However, the same point can be illustrated by assuming that the ranking ||IDENT[place] o *sTvT|| is fixed (since *state* is always pronounced as [steɪt]). The only difference is then that the prediction is that we will observe the ranking ||IDENT[place] o *sPvP|| 22% of the time. We are therefore expecting that an /sPvP/-input should map faithfully onto itself at least 22% of the time. An impossible word is now predicted to be possible at least some of the time.

I have illustrated the point here with a specific example. However, this is a general problem that stochastic grammars will face whenever possible and impossible words are compared.

4.2.3 Crucially unranked constraints

Anttila (1997) proposes an extension to the classic OT grammar in order to account for non-categorical (variable) phenomena. He assumes that the constraint hierarchy for some language can contain a set of crucially unranked constraints. On every evaluation occasion one complete ranking is selected at random from amongst the different complete rankings possible between the crucially unranked constraints. Variation arises if the different rankings between these crucially unranked constraints select different candidates as optimal. This theory also makes explicit predictions about absolute frequencies. Suppose that there are n unranked constraints. There are then $n!$ possible complete rankings between these constraints. Now suppose that m of these $n!$ possible rankings select some candidate as optimal. The prediction is that this candidate would be selected as optimal $m/n!$ of the time.[72]

How could this model be applied to the well-formedness rating and lexical decision experiments discussed above? Consider the comparative well-formedness judgment experiment for English (see §3.2.1.2) as an example. In this experiment subjects had to compare *inter alia* [sTvT]-forms with [sPvP]-forms. On some of these

---

[72] Independently from Anttila, Reynolds proposed a very similar extension to classic OT (Reynolds, 1994). Rather than having a set of crucially unranked constraints, Reynolds assumes that there are "floating constraints". A floating constraint is a constraint that is unranked relative a span of the (ranked) constraint hierarchy. This floating constraint version of OT can be applied to well-formedness judgments and lexical decisions in the same way as Anttila's crucially unranked constraint theory. I therefore do not discuss the floating constraint theory separately.

comparisons they preferred [sTvT] over [sPvP] (78% of the time), but on other comparison occasions they preferred [sPvP] over [sTvT] (22% of the times) (see (74b) above). The well-formedness rating |sTvT ™ sPvP| would result from the ranking ||*sPvP o *sTvT||, and the rating |sPvP ™ sTvT| would result from the ranking ||*sTvT o *sPvP||. The fact that both |sTvT ™ sPvP| and |sPvP ™ sTvT| were observed can be interpreted as evidence that the constraints *sTvT and *sPvP are crucially unranked. Every time a subject in the experiment had to compare an [sPvP]-form and an [sTvT]-form, one of the two rankings between these two constraints is selected at random. If the ranking ||*sPvP o *sTvT|| is selected, then the subject prefers [sTvT] over [sPvP]. If the ranking ||*sTvT o *sPvP|| is selected, the subject prefers [sPvP] over [sTvT].[73]

This model seems capable of accounting for the variation in the responses of the subjects. However, this account faces the same three problems as the stochastic account discussed just above in §4.2.2. In addition to these, it is also faced by another more practical problem. I will discuss this practical problem first, and then briefly show why this model faces the same problem as the stochastic models.

A practical problem. In the comparison between [sTvT] and [sPvP], [sTvT] was preferred more often. Assume that *sTvT and *sPvP are indeed unranked with regard to each other. There are only two possible rankings between these two constraints. If it is indeed the case that one of the possible rankings between the unranked constraints are selected at random at every evaluation occasion, then the ranking ||*sTvT o *sPvP|| will

---

[73] Of course, subjects also compared [sTvT]-forms with [sKvK]-forms, and [sKvK]-forms with [sPvP]-forms. Also in these comparisons the preferences went in both directions. The implication would therefore be that all three of the constraints *sTvT, *sKvK and *sPvP are crucially unranked with respect to each other.

be selected half the time and the opposite ||*sPvP o *sTvT|| half the time. From this follows the prediction that [sTvT] will be preferred over [sPvP] half the time, and [sPvP] over [sTvT] half the time. And this is not what was observed – see that data in (74).

This problem stems from the fact that there is one constraint each against [sPvP]-forms and [sTvT]-forms. Had there been more constraints against [sPvP]-forms than against [sTvT]-forms, then it would be possible to derive the fact that [sTvT] is preferred more frequently. This problem is therefore not a general problem for the Anttila-theory. It is at least in principle possible to analyze the restriction on [sCvC]-forms such that an [sPvP]-form will violate more constraints than an [sTvT]-form.[74]

The which-value problem. Like the stochastic models, the crucially unranked constraint grammars also model absolute difference in well-formedness between different token types. Like the stochastic models, these grammars are therefore faced with the problem that there are several absolute values that measure that well-formedness difference between the token types. Different experiments show different absolute values, different individuals differ from each other. Which of these values should be modeled?

The grammar-only problem. Again like the stochastic grammars, these grammars model all variation in the response data via the grammar. This implies that grammar alone is responsible for the performance of subjects in the experiments. It does not leave enough room for others factors that are known to influence performance – factors such as lexical statistics, experimental design, individual differences, etc.

---

[74] For instance, it is possible that the *sCvC-constraints stand to each other in a stringency relation – i.e. [sPvP] violates all three of the *sCvC-constraints, [sKvK] violates *sKvK and *sTvT, and [sTvT] violates only *sTvT. See de Lacy (2002, 2003) for such an interpretation of markedness scales.

Since the rank-ordering model of EVAL models only relative well-formedness differences between token types it can avoid both of these problems. It does not have to select a specific set of absolute values. And it does not model all aspects of the variation. It only stipulates the relative well-formedness differences between tokens. The absolute differences are determined by a complex interaction of grammar with many non-grammatical factors.

Relative well-formedness differences between possible and impossible words. This exact same problem is faced by stochastic grammars, and it has to do with the fact that possible words are not rated infinitely better than impossible words. Since [sPvP] is never a possible word, the ranking ||*sPvP o IDENT[place]|| is fixed. /sPvP/ is then always mapped unfaithfully onto something like [sPvT] or [sKvP]. In the comparative well-formedness judgment experiment [sPvP] was preferred over [sTvT] 22% of the time (see table (74b) above). This means that the constraints *sPvP and *sTvT should be allowed to freely rerank. At least some of the time the ranking ||*sTvT o *SPvP|| will then be observed. But because of the fixed ranking ||*sPvP o IDENT[place]||, this means that the ranking ||*sTvT o IDENT[place]|| will also be observed at least some of the time. And from this follows the prediction that an /sTvT/-input should some times not be allowed to map faithfully onto itself. It is predicted that the word *state* should some times be pronounced as something like [steɪt] or [steɪk].[75]

---

[75] I have sketched the argument here assuming that the ranking ||*sPvP o IDENT[place]|| is fixed. However, the same point can be illustrated by assuming that the ranking ||IDENT[place] o *sTvT|| is fixed (since *state* is always pronounced as [steɪt]). The only difference is then that the prediction is that we will observe the ranking ||IDENT[place] o *sPvP|| some of the time. We are therefore expecting that an /sPvP/-input should map faithfully onto itself at least some of the time. An impossible word is now predicted to be possible at least some of the time.

This problem was illustrated here with a specific example. However, this is a problem that will arise every time a possible and an impossible word are compared.

The rank-ordering model of EVAL avoids all of these problems. In this model non-categorical behavior of language users does not arise from variation in the grammar (in the constraint ranking). For English the constraints are always ranked ||*sPvP o *sKvK o IDENT[place] o *sTvT o Cut-off||. Since the ranking between the constraints does not vary, IDENT[place] always outranks *sTvT so that [sTvT] is always predicted as a possible word. Similarly, both *sPvP and *sKvK always outrank IDENT[place], so that neither [sKvK] nor [sPvP] is every predicted as a possible word.

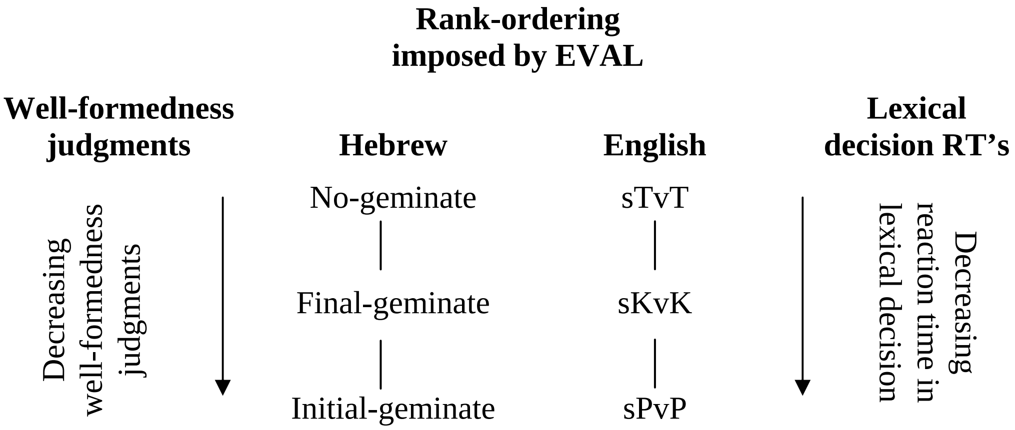
5. Concluding remarks

In this dissertation I argue for two extensions to a classic OT grammar: (i) First, I argue that EVAL should be allowed to compare forms that are not related to each other via a shared input. (ii) Secondly, I argue that EVAL imposes a harmonic rank-ordering on the full candidate set. In this chapter I have shown how these two extensions enable us to account for response patterns observed in well-formedness judgment experiments and lexical decision experiments.

In well-formedness judgment experiments and lexical decision experiments language users are asked to compare different non-words (either directly as in comparative well-formedness experiments, or implicitly as in gradient well-formedness judgment experiments and lexical decision experiments). Non-words used in these experiments are not related to each other via a shared input. The assumption that these forms can be grammatically compared therefore implies that EVAL can compare such morphologically unrelated forms.

I have argued that the influence of grammar on well-formedness judgments and lexical decision can be accounted for as follows: The more well-formed a token is according to grammar (the higher slot it occupies in the rank-ordering imposed by EVAL on the candidate set), the more well-formed it will judged to be. Similarly, a non-word that is more well-formed according to grammar will be considered more seriously as potential word and will therefore be rejected more slowly as a non-word in a lexical decision task. I have illustrated this with an example on how the OCP influences the processing of non-words in Hebrew, and how a restriction on [sCvC]-forms influences the processing of non-words in English. The predictions about Hebrew and English are represented graphically in (76).

(76) **Hebrew and English in a rank-ordering model or EVAL**



The results of the experiments discussed above confirmed these predictions. This is an important result for two reasons: First, it serves as evidence for the rank-ordering model of EVAL – the predictions that follow from this model are confirmed. Secondly, it also serves as evidence for the analyses of Hebrew geminates and of English [sCvC]-

forms. The fact that the subjects in the experiments responded according to the predictions that follow from the analyses, confirms the correctness of the analyses.

The rank-ordering model of EVAL then serves two purposes: It expands the coverage of our theory in that we can now account for non-categorical phenomena such as well-formedness judgments and reaction times in lexical decision. Secondly, it adds to the kind of data that we can use as information about the grammar of some language – we can now use the results of well-formedness judgment experiments and lexical decision experiments as data to develop and test grammatical analyses.

Appendix: Tokens used in [sCvC]-experiments

A.1 Gradient well-formedness judgment experiment (see §3.2.1.1)

(77) **[sTvT] and [sKvK]-tokens**

[sTvT]	ND	Cumulative Probability	[sKvK]	ND	Cumulative Probability
stɔɪt	10.65	6.92×10^{-12}	skeɪk	26.06	2.24×10^{-9}
stɑːt	44.16	2.56×10^{-8}	skæk	22.12	6.45×10^{-9}
stɔːt	23.06	3.9×10^{-9}	skɑːk	18.41	2.26×10^{-9}
stʊt	17.06	4.36×10^{-10}	skɛk	12.65	2.41×10^{-9}
stʌt	42.19	1.43×10^{-8}	skɪk	37.31	2.16×10^{-8}
	27.43	8.85×10^{-9}		23.31	6.99×10^{-9}

***t*-test on difference in ND:** *t*(8) = 0.52, two-tailed *p* = 0.62.
***t*-test on difference in Probability:** *t*(8) = 0.30, two-tailed *p* = 0.77.

(78) **[sTvT] and [sPvP]-tokens**

[sTvT]	ND	Cumulative Probability	[sPvP]	ND	Cumulative Probability
stʊt	17.06	4.36×10^{-10}	spɑːp	21.41	2.84×10^{-9}
stɔɪt	10.65	6.92×10^{-12}	spɛp	20.32	2.40×10^{-9}
stʌt	42.19	1.43×10^{-8}	spɪp	26.69	5.48×10^{-9}
stɑːt	44.16	2.56×10^{-8}	spæp	17.98	2.24×10^{-9}
stɔːt	23.06	3.9×10^{-9}	spiːp	27.45	4.23×10^{-9}
	27.43	8.85×10^{-9}		22.77	3.44×10^{-9}

***t*-test on difference in ND:** *t*(8) = 0.67, two-tailed *p* = 0.52.
***t*-test on difference in Probability:** *t*(8) = 1.09, two-tailed *p* = 0.31.

(79) **[sKvK] and [sPvP]-tokens**

[sKvK]	ND	Cumulative Probability	[sPvP]	ND	Cumulative Probability
skaɪk	26.06	2.24×10^{-9}	spæp	17.98	2.24×10^{-9}
skaʊk	9.95	2.45×10^{-14}	spɑːp	21.41	2.84×10^{-9}
skiːk	28.63	5.35×10^{-10}	spɛp	20.32	2.40×10^{-9}
skɔːk	10.04	3.88×10^{-10}	spɪp	26.69	5.48×10^{-9}
skuːk	12.64	6.30×10^{-11}	spiːp	27.45	4.23×10^{-9}
	17.46	6.45×10^{-10}		22.77	3.44×10^{-9}

***t*-test on difference in ND:** *t*(8) = 1.18, two-tailed *p* = 0.27.
***t*-test on difference in Probability:** *t*(8) = 3.76, one-tailed *p* = 0.005.

A.2 Comparative well-formedness judgment experiments (see §3.2.1.2)

(80) [sTvT]~[sKvK]

[sTvT]	ND	Cumulative Probability	[sKvK]	ND	Cumulative Probability	Diff in ND	Diff in Cumulative Probability
stɔ:t	23.06	3.90×10^{-9}	skæk	22.12	6.45×10^{-9}	0.95	-2.55×10^{-9}
stʊt	17.06	4.36×10^{-10}	skɑ:k	18.41	2.26×10^{-9}	-1.35	-1.83×10^{-9}
stɔɪt	10.65	6.92×10^{-12}	skɛk	12.65	2.41×10^{-9}	-2.00	-2.41×10^{-9}
stɔɪt	10.65	6.92×10^{-12}	skɑ:k	18.41	2.26×10^{-9}	-7.76	-2.26×10^{-9}
stɔɪt	10.65	6.92×10^{-12}	skʌk	19.38	1.92×10^{-9}	-8.73	-1.91×10^{-9}
stɔɪt	10.65	6.92×10^{-12}	skæk	22.12	6.45×10^{-9}	-11.47	-6.44×10^{-9}
stʊt	17.06	4.36×10^{-10}	skeɪk	26.06	2.24×10^{-9}	-9.00	-1.81×10^{-9}
stʌt	42.19	1.43×10^{-8}	skɪk	37.31	2.16×10^{-8}	4.88	-7.27×10^{-9}
stɔ:t	23.06	3.90×10^{-9}	skɪk	37.31	2.16×10^{-8}	-14.25	-1.77×10^{-8}
stʊt	17.06	4.36×10^{-10}	skɪk	37.31	2.16×10^{-8}	-20.25	-2.12×10^{-8}
stʊt	17.06	4.36×10^{-10}	skæk	22.12	6.45×10^{-9}	-5.05	-6.01×10^{-9}
stɔɪt	10.65	6.92×10^{-12}	skaɪk	13.89	6.67×10^{-10}	-3.24	-6.60×10^{-10}
stɔɪt	10.65	6.92×10^{-12}	skeɪk	26.06	2.24×10^{-9}	-15.41	-2.24×10^{-9}
stʊt	17.06	4.36×10^{-10}	skɛk	12.65	2.41×10^{-9}	4.42	-1.98×10^{-9}
stɔɪt	10.65	6.92×10^{-12}	skɪk	37.31	2.16×10^{-8}	-26.66	-2.16×10^{-8}
	16.54	1.62×10^{-9}		24.21	8.14×10^{-9}		

***t*-test on difference in ND:** $t(14) = 3.32$, one-tailed $p = 0.003$.
***t*-test on difference in Probability:** $t(14) = 3.43$, one-tailed $p = 0.002$.

(81) [sTvT]~[sPvP]

[sTvT]	ND	Cumulative Probability	[sPvP]	ND	Cumulative Probability	Diff in ND	Diff in Cumulative Probability
stɔ:t	23.06	3.9×10^{-9}	spɪp	26.69	5.48×10^{-9}	-3.63	-1.58×10^{-9}
stɔɪt	10.65	6.9×10^{-12}	spʌp	11.38	1.31×10^{-9}	-0.73	-1.31×10^{-9}
stɔɪt	10.65	6.9×10^{-12}	spi:p	27.45	4.23×10^{-9}	-16.80	-4.23×10^{-9}
stɔɪt	10.65	6.9×10^{-12}	spɪp	26.69	5.48×10^{-9}	-16.04	-5.47×10^{-9}
stɔ:t	23.06	3.9×10^{-9}	spi:p	27.45	4.23×10^{-9}	-4.39	-3.31×10^{-10}
stɔɪt	10.65	6.9×10^{-12}	spɑ:p	21.41	2.84×10^{-9}	-10.76	-2.83×10^{-9}
stɔɪt	10.65	6.9×10^{-12}	spæp	17.98	2.24×10^{-9}	-7.33	-2.23×10^{-9}
stʊt	17.06	4.4×10^{-10}	spæp	17.98	2.24×10^{-9}	-0.92	-1.80×10^{-9}
stʊt	17.06	4.4×10^{-10}	spɪp	26.69	5.48×10^{-9}	-9.62	-5.04×10^{-9}
stɔɪt	10.65	6.9×10^{-12}	speɪp	17.13	8.52×10^{-10}	-6.48	-8.45×10^{-10}
stɔɪt	10.65	6.9×10^{-12}	spɛp	20.32	2.40×10^{-9}	-9.67	-2.40×10^{-9}
stʊt	17.06	4.4×10^{-10}	spɑ:p	21.41	2.84×10^{-9}	-4.35	-2.40×10^{-9}
stʊt	17.06	4.4×10^{-10}	spi:p	27.45	4.23×10^{-9}	-10.39	-3.80×10^{-9}
stʊt	17.06	4.4×10^{-10}	spɛp	20.32	2.40×10^{-9}	-3.25	-1.97×10^{-9}
stʊt	17.06	4.4×10^{-10}	spʌp	11.38	1.31×10^{-9}	5.68	-8.76×10^{-10}
	14.87	6.99×10^{-10}		21.45	3.17×10^{-9}		

***t*-test on difference in ND:** $t(14) = 4.30$, one-tailed $p < 0.000$.
***t*-test on difference in Probability:** $t(14) = 6.21$, one-tailed $p < 0.000$.

(82) **[sKvK]~[sPvP]-tokens**

[sKvK]	ND	Cumulative Probability	[sPvP]	ND	Cumulative Probability	Diff in ND	Diff in Cumulative Probability
skaʊk	9.95	2.45×10^{-14}	spæp	17.98	2.24×10^{-9}	-8.03	-2.24×10^{-9}
sku:k	12.64	6.30×10^{-11}	spæp	17.98	2.24×10^{-9}	-5.34	-2.18×10^{-9}
skaƱk	9.95	2.45×10^{-14}	spɑ:p	21.41	2.84×10^{-9}	-11.46	-2.84×10^{-9}
sku:k	12.64	6.30×10^{-11}	spɑ:p	21.41	2.84×10^{-9}	-8.77	-2.78×10^{-9}
skʊk	3.86	5.76×10^{-10}	spɑ:p	21.41	2.84×10^{-9}	-17.55	-2.26×10^{-9}
skaʊk	9.95	2.45×10^{-14}	spɛp	20.32	2.4×10^{-9}	-10.37	-2.40×10^{-9}
sku:k	12.64	6.30×10^{-11}	spɛp	20.32	2.4×10^{-9}	-7.68	-2.34×10^{-9}
skʊk	3.86	5.76×10^{-10}	spɛp	20.32	2.4×10^{-9}	-16.46	-1.83×10^{-9}
ski:k	28.63	5.35×10^{-10}	spɪp	26.69	5.48×10^{-9}	1.94	-4.95×10^{-9}
skɔ:k	10.04	3.88×10^{-10}	spɪp	26.69	5.48×10^{-9}	-16.64	-5.09×10^{-9}
skʊk	3.86	5.76×10^{-10}	spɪp	26.69	5.48×10^{-9}	-22.83	-4.90×10^{-9}
skɔ:k	10.04	3.88×10^{-10}	spi:p	27.45	4.23×10^{-9}	-17.41	-3.85×10^{-9}
skʊk	3.86	5.76×10^{-10}	spi:p	27.45	4.23×10^{-9}	-23.59	-3.66×10^{-9}
skaʊk	9.95	2.45×10^{-14}	spʌp	11.38	1.31×10^{-9}	-1.43	-1.31×10^{-9}
sku:k	12.64	6.30×10^{-11}	spʌp	11.38	1.31×10^{-9}	1.26	-1.25×10^{-9}
	10.30	2.58×10^{-10}		21.26	3.18×10^{-9}		

t-test on difference in ND: $t(14) = 5.26$, one-tailed $p < 0.000$.
t-test on difference in Probability: $t(14) = 8.87$, one-tailed $p < 0.000$.

A.3 Lexical decision experiment (see §3.2.2)

(83) **[sTvT] and [sKvK]-tokens**

[sTvT]	ND	Cumulative Probability	[sKvK]	ND	Cumulative Probability
stɔ:t	23.06	3.90×10^{-9}	skeɪk	26.06	2.24×10^{-9}
staɪt	40.01	5.96×10^{-8}	ski:k	28.63	5.35×10^{-10}
stɛt	38.11	5.16×10^{-8}	skɪk	37.31	2.16×10^{-8}
stu:t	39.66	1.27×10^{-7}	skæk	13.89	6.45×10^{-9}
stʊt	17.06	4.36×10^{-10}	sku:k	22.12	6.30×10^{-11}
	31.58	4.85×10^{-8}		25.60	6.18×10^{-9}

t-test on difference in ND: $t(8) = 0.97$, two-tailed $p = 0.36$.
t-test on difference in Probability: $t(8) = 1.81$, two-tailed $p = 0.11$.

(84) **[sTvT] and [sPvP]-tokens**

[sTvT]	**ND**	**Cumulative Probability**	**[sPvP]**	**ND**	**Cumulative Probability**
stɔ:t	23.06	3.90×10^{-9}	spɑ:p	21.41	2.84×10^{-9}
staɪt	40.01	5.96×10^{-8}	spaɪp	30.94	9.92×10^{-11}
stɛt	38.11	5.16×10^{-8}	spi:p	27.45	4.23×10^{-9}
stu:t	39.66	1.27×10^{-7}	spɪp	26.69	5.48×10^{-9}
stʊt	17.06	4.36×10^{-10}	spu:p	26.28	4.64×10^{-11}
	31.58	4.85×10^{-8}		26.56	2.54×10^{-9}

t-test on difference in ND: $t(8) = 0.99$, two-tailed $p = 0.35$.
t-test on difference in Probability: $t(8) = 1.99$, two-tailed $p = 0.08$.

(85) **[sKvK] and [sPvP]-tokens**

[sKvK]	**ND**	**Cumulative Probability**	**[sPvP]**	**ND**	**Cumulative Probability**
skaɪk	13.89	6.67×10^{-10}	spaɪp	30.94	9.92×10^{-11}
skaʊk	9.95	2.45×10^{-14}	spaʊp	5.20	1.30×10^{-13}
skɑ:k	18.41	2.26×10^{-9}	spɑ:p	21.41	2.84×10^{-9}
ski:k	28.63	5.35×10^{-10}	spɔ:p	2.38	9.92×10^{-13}
skʊk	3.86	5.77×10^{-10}	spʌp	11.38	1.31×10^{-9}
	14.95	8.08×10^{-10}		14.26	8.50×10^{-10}

t-test on difference in ND: $t(8) = 0.10$, two-tailed $p = 0.92$.
t-test on difference in Probability: $t(8) = 0.06$, two-tailed $p = 0.96$.