Grammar is both categorical and gradient[1]

Andries W. Coetzee

Abstract. In this paper, I discuss the results of word-likeness rating experiments with Hebrew and English speakers that show that language users use their grammar in a categorical and a gradient manner. In word-likeness rating tasks, subjects make the categorical distinction between grammatical and ungrammatical – they assign all grammatical forms equally high ratings and all ungrammatical forms equally low ratings. However, in comparative word-likeness tasks, subjects are forced to make distinctions between different grammatical or ungrammatical forms. In these experiments, they make finer gradient well-formedness distinctions. This poses a challenge on the one hand to standard derivational models of generative grammar, which can easily account for the categorical distinction between grammatical and ungrammatical, but have more difficulty with the gradient well-formedness distinctions. It also challenges models in which the categorical distinction between grammatical and ungrammatical does not exist, but in which an ungrammatical form is simply a form with very low probability. I show that the inherent comparative character of an OT grammar enables it to model both kinds of behaviors in a straightforward manner.

Introduction

There is a growing body of literature showing that phonological grammar influences phonological performance. We know that grammar plays a role in phoneme identification (Coetzee, 2005; Massaro and Cohen, 1983; Moreton, 2002), the segmentation of speech into words (Kirk, 2001; Suomi *et al*., 1997), lexical decision (Berent, Shimron and Vaknin, 2001), word-likeness ratings (Berent, Everett and Shimron, 2001; Frisch and Zawaydeh, 2001), etc. Once we accept that performance reflects the influence of grammar, we can use performance data as a window on what grammar looks like. In this paper, I discuss performance data showing that grammar is categorical and gradient. Grammar must be able to distinguish between grammatical (possible words) and ungrammatical (impossible words). However, grammar must also be able to make gradient well-formedness distinctions within these two sets. In the set of grammatical forms, there are some forms that are "more" and some that are "less" grammatical. Similarly, there are "more" and "less" ungrammatical forms.[2]

These data speak to the very core of grammar. They show that standard generative models in which grammar is simply a function that maps every input onto its unique grammatical output cannot be entirely correct. This would be equivalent to a grammar that makes only the categorical grammatical/ungrammatical distinction. On the other hand, it also shows that models in which grammaticality is only a value on a continuous scale of probability cannot be correct. We need a model of grammar that can make both the qualitative, categorical distinction between grammatical and ungrammatical, and gradient distinctions within the sets of grammatical and ungrammatical forms. I will show that the connections of Optimality Theory (OT) to standard generative grammar enable OT to draw the distinction between grammatical and ungrammatical in a

1

straightforward manner. However, because of its inherent comparative nature it can also easily model gradient distinctions in well-formedness.

This paper is structured as follows: I start out with a general discussion of the relationship between grammar and word-likeness judgments. The next section discusses the results of word-likeness experiments performed by Berent and colleagues (Berent and Shimron, 1997; Berent, Everett and Shimron, 2001; Berent, Shimron and Vaknin, 2001) with Hebrew speakers. These experiments show that Hebrew speakers categorically distinguish between grammatical and ungrammatical forms in some task conditions, but that they also make gradient well-formedness distinctions in other task conditions. After discussion of the Hebrew experiments, I discuss similar experiments that I conducted with English listeners. These experiments confirm the results of Berent's Hebrew experiments. Once the experimental results have been presented, I develop a straightforward way in which to account for these results within OT. Finally, I show why the results of the experiments are problematic for other grammatical models.

Grammar and word-likeness judgments

It is well known that language users have strong intuitions about what counts as a possible word of their language. Although [blɪk] is not an actual word of English, it is perfectly well-formed according the phonotactics of English. *[knɪk], on the other hand, violates a phonotactic constraint – English does not tolerate [#kn-] word-initially (Chomsky and Halle 1965:101). If these two forms were presented to English speakers in a word-likeness rating task, [blɪk] would receive higher ratings than *[knɪk]. This can be interpreted as evidence for the influence of grammar on word-likeness ratings. However, there is a confound that sheds doubt on this interpretation. Nonce words that contain phoneme sequences that occur frequently in the lexicon are rated as more word-like than nonce words with less frequent phoneme sequences (Bailey and Hahn, 1998; Hay et al., 2004; Coleman and Pierrehumbert, 1997; etc.). Low ratings for nonce words with phonotactically illegal (and therefore non-occurring) sequences can then be interpreted as the logical extreme of such a frequency bias – a better rating for [blɪk] might simply reflect the fact that [#bl-] has a higher frequency than [#kn-] in the English lexicon.

But there is also experimental evidence showing that grammar does contribute independently from this kind of frequency statistics in the lexicon to phonological processing. In a study of word-likeness ratings in Arabic, Frisch and Zawaydeh (2001) used non-words containing unattested consonant sequences. Half of their stimuli contained consonant sequences that were absent from the lexicon because they violated a systematic phonotactic constraint of Arabic (they contained contiguous homorganic consonants in violation of the Obligatory Contour Principle). The other half contained sequences that they characterize as accidental gaps since none of the sequences belong to a coherent natural class of non-occurring consonant pairs (i.e. there are no phonotactic constraints against them). Both of these kinds of tokens contain non-

2

occurring sequences and therefore do not differ in terms of the phoneme sequence frequency statistics calculated over the lexicon. However, they found that tokens that contained OCP-violating sequences received lower word-likeness ratings than other tokens. Since both of these token types contain non-occurring sequences, this difference cannot originate in lexical statistics. They ascribe the difference to grammar.

Results such as these support the hypothesis that grammar plays a role in word-likeness judgments. If we accept this hypothesis as true, we can look at word-likeness judgments for information on the structure of grammar – this is the topic of the next two sections. I discuss two kinds of word-likeness judgment tasks. Language users employ the information provided by grammar differently in the two tasks, and consequently treat the same (kind of) token differently. In a "word-likeness rating" experiment, subjects are presented with one nonce word token at a time, and they have to assign each token a rating from some rating scale. In a "comparative word-likeness" experiment, subjects are presented with more than one token at a time, and they have to order the tokens according to their word-likeness.

In the experiments that I discuss below, we find evidence for the categorical nature of grammar in the word-likeness rating experiments. All nonce words that are well-formed according to the grammar received relatively high ratings. Consequently, these tokens are not distinguished from each other in terms of their assigned ratings. Similarly, all nonce words that are phonotactically ill-formed received very low ratings and were not distinguished from each other. Although the subjects had a rating scale with several discrete values, they treated the task as an "accept or reject" task, using basically only two values on the scale. Language users can therefore use the information provided by grammar to make a categorical distinction between grammatical and ungrammatical.

In the comparative word-likeness experiments, we find evidence for the gradient nature of grammar. In these experiments, subjects are sometimes required to compare two grammatical nonce words or two ungrammatical nonce words with each other, and to select the one that is more word-like. In the word-likeness rating task, they might have assigned two grammatical nonce words equally high ratings. But now that option is not available, and we find the following: although two nonce words might both be grammatical, it is possible that one contains a more marked structure and is therefore less well-formed. When forced to choose between two such forms, language users prefer the more well-formed token. The same happens when they are forced to compare two ungrammatical forms. Two forms might both be ungrammatical because they contain marked structures not tolerated in the language. However, one of the forms might contain a more marked structure. When forced to choose between two such nonce words, language users prefer the one that is less ill-formed. In addition to the categorical grammatical/ungrammatical distinction, language users can also make finer gradient distinctions in terms well-formedness.

In the next two sections, I discuss word-likeness experiments in Hebrew and English that illustrate these two uses of the information provided by grammar. Once the results of the experiments have been presented, I develop a version of OT that can account for the two response strategies.

3

Word-likeness ratings in Hebrew

One of the most striking features of Semitic morphology is the limitation on the distribution of identical consonants in verbal roots (Frisch *et al*., 2004; Gafos, 2003; Greenberg, 1950; McCarthy, 1986, 1994; Pierrehumbert, 1993; etc.). Forms with identical initial consonants are not allowed – *[X-X-Y] is ill-formed. On the other hand, forms with identical final consonants are well-formed – [X-Y-Y] is acceptable.[3] I will refer to *[X-X-Y]-forms as "initial-geminates", to [X-Y-Y]-forms as "final-geminates", and to forms with no identical consonants, e.g. [X-Y-Z]-forms, as "non-geminates".

Berent and colleagues (Berent, Everett and Shimron, 2001; Berent, Shimron and Vaknin, 2001; Berent and Shimron, 1997) conducted a series of experiments in which they tested whether this restriction plays a role in how Hebrew speakers rate nonce words. In word-likeness rating tasks, they found that Hebrew speakers rated the two kinds of possible words, [X-Y-Z] and [X-Y-Y], equally good and both better than the ungrammatical *[X-X-Y]-forms. However, in comparative word-likeness tasks, their subjects differentiated between the two kinds of grammatical tokens – they preferred the non-geminates over the final-geminates. Although both of these are grammatical, the final geminates contain a marked structure (geminate consonants) absent from the non-geminates. When forced to choose between them, subjects go for the less marked token. I discuss the experiments of Berent and Shimron (1997) as a representative example of these experiments.

Word-likeness rating[4]

Berent and Shimron (1997) selected 24 root trios. None of the roots corresponded to an existing Hebrew word. One of the members in each trio had three non-identical consonants (henceforth referred to as the "non-geminate" member). The other members both shared the first two consonants of the non-geminate. One of them doubled the first consonant forming an initial-geminate, and the other doubled the second consonant forming a final-geminate. Each trio had the structure [X-X-Y]~[X-Y-Y]~[X-Y-Z]. Each of these trios was conjugated in three verbal forms. Their stimuli therefore included 72 non-geminate nonce words (24 non-geminate roots conjugated in 3 verbal forms), 72 final-geminate and 72 initial-geminate nonce words. The tokens were randomized, and presented in a written word-likeness rating task to 15 native speakers of Hebrew, all of whom were psychology students at Haifa University in Israel. Subjects had to rate each token on a 5-point scale, with [1] corresponding to a form that is impossible as a word of Hebrew and [5] to a form that is an excellent candidate for a Hebrew word.

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.[5] The results of this experiment are summarized in (1), and represented graphically in Figure 1.

(1)   Difference scores in word-likeness rating experiment

Comparison	Example	Difference score	t	df	p
Initial-geminates and non-geminates	*[X-X-Y] [X-Y-Z]	0.881	11.1	46	< 0.001
Initial-geminates and final-geminates	*[X-X-Y] [X-Y-Y]	0.801	10.0	46	< 0.001
Final-geminates and non-geminates	[X-Y-Y] [X-Y-Z]	0.081	_[6]		> 0.05

Figure 1: Difference scores in the word-likeness experiment of Berent and Shimron (1997)
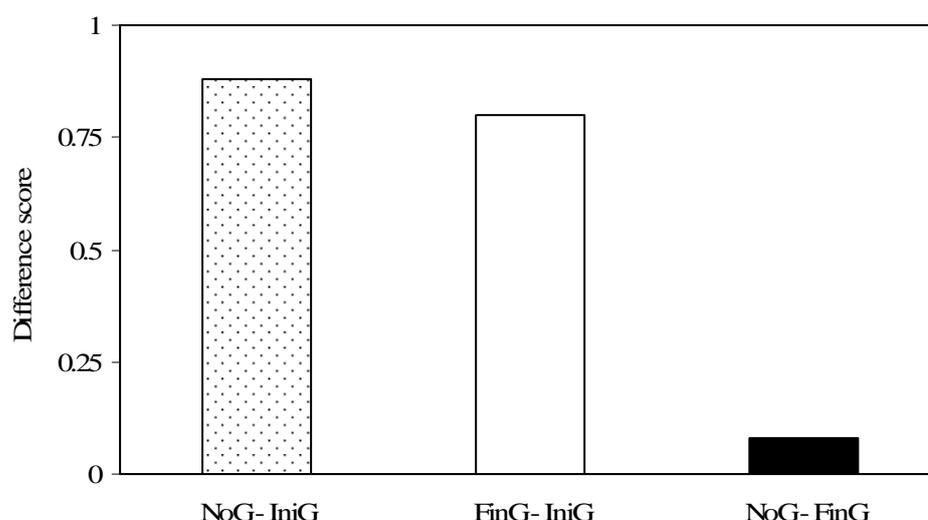

Comparative word-likeness

In a second experiment, Berent and Shimron used the same 24 root trios conjugated in the same three verbal patterns as in the word-likeness rating experiment. They therefore had 72 non-word trios (24 trios conjugated in 3 verbal forms), each trio with a non-geminate, a final-geminate and an initial-geminate. Unlike in the first experiment, where the forms were presented one at a time, the three members of each trio were presented together in the comparative word-likeness experiment. The order among the members of each trio was randomized, and the trios themselves were also randomized. These trios were presented in written form to 18 students from Haifa University, all of whom were native speakers of Hebrew. Their task was to order the members of each trio in terms of its word-likeness. A score of [3] was assigned to the most word-like member, and a score of [1] to the least word-like member. This setup differs from the word-likeness rating experiment by forcing subjects to choose between the two kinds of possible words.

As with the word-likeness rating experiment, Berent and Shimron report only difference scores. Their results are summarized in (2) and represented graphically in Figure 2.

5

(2)   Difference scores in comparative word-likeness experiment

Comparison	Example	Difference score	t	df	p
Initial-gemination and no-gemination	*[X-X-Y] [X-Y-Z]	1.122	18.6	46	< 0.001
Initial-gemination and final-gemination	*[X-X-Y] [X-Y-Y]	0.682	11.3	46	< 0.001
Final-gemination and no-gemination	[X-Y-Y] [X-Y-Z]	0.44	_[7]		< 0.05

Figure 2: Difference scores in the comparative word-likeness experiment of Berent and Shimron (1997)
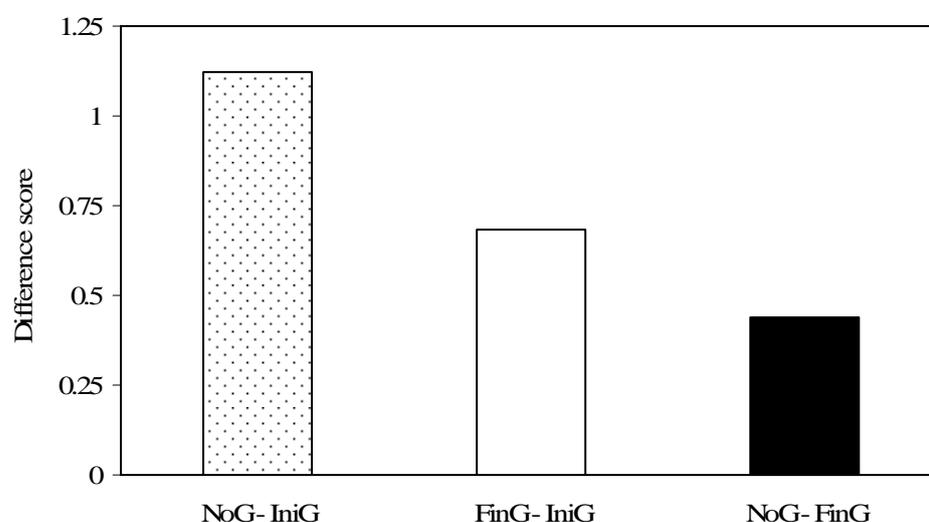
In both experiments, their subjects differentiate initial-geminates from non-geminates and final-geminates. This shows that (i) there is a difference between these token types in well-formedness, and (ii) both kinds of tasks are sensitive enough to pick up on this difference. In the comparative word-likeness rating task, the subjects also distinguished between non-geminates and final-geminates, so we know that these two token types also differ in terms of their well-formedness. (See below for the reasons for this difference in well-formedness.) What is interesting is that the subjects did not distinguish between these two token types in the word-likeness rating task, and this in spite of the fact that the word-likeness rating task can detect differences in well-formedness, and that there is a difference in well-formedness between non-geminates and final-geminates. Inspection of Figure 1 will show that there is a small difference between the non-geminates and the final-geminates even in the word-likeness rating task, and that this difference is in the same direction as in the comparative word-likeness rating task. It is possible that this difference between non-geminates and final-geminates would also have reached significance if more data were collected. However, even if that were the case it would still be true that the subjects treat the difference between the ungrammatical tokens (initial-geminates) and grammatical tokens (non-geminates and final-geminates) the same in the two tasks, but they treat the difference

6

between the two kinds of grammatical tokens (non-geminates and final-geminates) differently between the two tasks. The latter difference is significantly decreased in the word-likeness rating task while the former is not.

The results of these two experiments show that language users can use the information provided by grammar in two different ways. In some tasks, they use the information in a categorical manner to distinguish grammatical forms from ungrammatical forms. In other tasks, they use grammar to make finer gradient well-formedness distinctions between different grammatical forms.

Word-likeness ratings in English

English restricts the consonants that can co-occur in the onset and coda of a syllable (Fudge, 1969; Davis, 1984). I focus on only one aspect of this restriction here – words of the form [sCVC] are tolerated if both C's are [t], but not if both are [k] or [p] – *state* is a word, but words of the form *skake and *spape do not exist in English. Following Davis (1984, 1991), I will interpret the absence of *skake and *spape words as evidence that these are not possible words of English (see also Browne, 1981; Clements and Keyser, 1983; Fudge, 1969; Lamontagne, 1993: Chapter 6; etc.). We therefore have a situation very similar to the Hebrew example above – we have a phonotactic constraint that can be used to divide a set of non-words between grammatical forms ([stVt]) and ungrammatical forms (*[skVk] and *[spVp]). However, the situation is also different from the Hebrew example – in English we have two kinds of ungrammatical forms (*[skVk] and *[spVp]) rather than two kinds of grammatical forms as we had in Hebrew.

This makes for an interesting way in which to replicate some of Berent and Shimron's results, and to extend upon their results. If English speakers react the same way as Hebrew speakers, then we would expect them to rate all grammatical forms ([stVt]) high, and all ungrammatical forms (*[skVk] and *[spVp]) low in a word-likeness rating experiment – i.e. we do not expect to see a difference between the two kinds of ungrammatical forms. However, if there is a well-formedness difference between *[spVp]- and *[skVk]-forms, we do expect to see evidence for this difference in a comparative word-likeness experiment. We therefore need to answer the following question: If there is a difference in word-likeness between *[skVk] and *[spVp], which of these two forms will be more and which will be less word-like? There is nothing in the phonological grammar of English that speaks to this question directly. However, there are several pieces of secondary evidence all of which converge on the conclusion that *[spVp] is most likely less well-formed than *[skVk]. I will briefly mention the most important aspects of this evidence here. For a more detailed discussion, see Coetzee (2004:395--8, 403--6).

In general, English restricts the co-occurrence of labials more severely than the co-occurrence of dorsals. There are two kinds of evidence for this. First, there are certain contexts in which two dorsals can occur but two labials cannot: (i) English tolerates words of the form [skVg] but not of the form *[spVb] – e.g. *skag* but *spab. (ii)

7

English tolerates words of the form [skVXk] where [X] stands for a nasal or a liquid; however, words of the forms *[spVXp] are not tolerated – e.g. *skulk*, *skunk*, but *spulp*, *spump*. (iii) Similarly, English allows words of the form [skGVk] where [G] is a glide, but words of the form *[spGVp] are not tolerated – e.g. *squeak* but *spweep*, *spyeep*. The second kind of evidence is not about possible and impossible words, but rather about statistical tendencies in the English lexicon. Berkley (1994, 2000) counted the number of English words with two homorganic consonants separated by at most two segments (i.e. *pop*, *palm*, *king*, *skulk*, *state*, *tact*, etc.). She then calculated the number of such words that would have been expected had consonants combined randomly. The ratio of the observed frequency to the expected frequency (O/E) is an index of the degree of over or underrepresentation of each word type. Berkley found that words with two dorsals and words with two labials were both underrepresented (i.e. had O/E-values below 1). However, the O/E-ratio for labials (0.57) was lower than that for dorsals (0.71) This shows that the co-occurrence of labials is more restricted than that of dorsals. See also de Lacy (2002:173 ff.) for arguments that labials are universally more marked than dorsals.

If we assume that the preference for the co-occurrence of dorsals will transfer onto [sCVC]-forms, then, even though both *[skVk] and *[spVp] are ungrammatical, *[skVk] will be more well-formed than *[spVp]. If this is true and if English subjects respond in a manner similar to Hebrew subjects, then English subjects should prefer *[skVk]-forms over *[spVp]-forms in a comparative word-likeness experiment. I performed a series of experiments with speakers of American English to test these predictions. All the experiments were performed during 2003 at the University of Massachusetts. The rest of this section is dedicated to discussing the design and results of these experiments.

Word-likeness rating

Design

Subjects. Twenty native speakers of American English were recruited from the undergraduate population at the University of Massachusetts. Most of the subjects grew up in western Massachusetts and were therefore speakers of the same dialect. However, because of exposure to speakers of other dialects, both in their daily lives and on the television and radio, it was not possible to control for differences between subjects in terms of their exposure to different dialects of American English. None of the subjects reported any speech or hearing deficit. Subjects received credit in an introductory linguistics class for their participation.

Token selection. Tokens were selected in three conditions: (i) T~K: 5 non-words each of the form [stVt] and *[skVk]; (ii) T~P: 5 non-words each of the form [stVt] and *[spVp]; (iii) K~P: 5 non-words each of the form *[skVk] and *[spVp]. All tokens were selected to control for the possible influence of lexical statistics on word-likeness rating (see the discussion above). Two kinds of lexical statistics were calculated for each token: lexical neighborhood density (LND) and cumulative bi-phone probability

8

(CBP).[8] The tokens were selected such that the tokens in each condition did not differ in terms of these statistics. If subjects treat these token types differently, the difference can therefore not be ascribed to a difference in these lexical statistics. The actual tokens used and their lexical statistics are included in the appendix.

Recordings. All tokens were read in the frame sentence "John said _____ again to me." Tokens were read by a phonetically trained native female speaker of American English. The speaker was in her early twenties, and spoke standard Midwestern American English. She had a clear, native distinction between the vowels in *pin* and *pen*. Each token was recorded 4 times. Recordings were made in a soundproofed booth at the Phonetics Laboratory of the University of Massachusetts. All tokens were excised from the frame sentence, and a single instance of each token was selected for use in the experiment. The instance selected was judged to be the clearest example of the specific token. This judgment was based on impressionistic grounds. All selected instances had released final stops (the release was possible because the word following the token in the frame sentence started with a vowel). This was confirmed by inspection of waveforms and spectrograms of the tokens.

Procedure. There were a total of 24 test-tokens.[9] Each of these tokens was included twice in the stimulus list. To these tokens were added 77 non-word filler items,[10] so that the stimulus list contained a total of 125 tokens. The stimulus list was presented auditorily to subjects twice so that each test-token was presented four times. There was a break of roughly 5 minutes between presentations. The stimulus list was randomized differently on every 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, and a score of [5] to a token that was judged to be very well-formed/very likely to be included in the lexicon of English. After 5 seconds, the next token was presented. Before the list was presented the first time, 10 filler tokens were presented as practice trials.

Results and discussion

The design of the experiment allows for a comparison of the grammatical forms ([stVt]) with each of the two ungrammatical forms (*[skVk] and *[spVp]), and for a comparison of the two kinds of ungrammatical forms with each other. If the English subjects respond in the same way as the Hebrew subjects, then we would expect the grammatical [stVt]-forms to be rated better than the ungrammatical forms, but we would expect no difference between the two kinds of ungrammatical forms.

Each token was presented four times. The mean score that each subject assigned to each token was calculated. Statistical analyses were performed on these mean scores. The scores were subjected to a 2×3 ANOVA with hypothesized grammatical well-formedness (more well-formed~less well-formed) and condition (K~P, T~K, T~P) as independent variables. The main effects of well-formedness ($F(1, 594) = 109.9$, $p < 0.001$) and condition ($F(2, 594) = 43.3$, $p < .001$) were both significant, as well as the

9

interaction between well-formedness and condition ($F(2, 594) = 41.9$, $p < .001$). The contrasts between the more and less well-formed tokens in each condition were further investigated with one-tailed t-tests. Since three comparisons are made, the critical value for significance is taken to be 0.0167 to control for type 1 errors. These tests returned significant differences for the T~P-condition ($t(198) = 9.1$, $p < .001$) and T~K-condition ($t(198) = 9.2$, $p < .001$), but not for the K~P-condition ($t(198) = 0.9$, $p = .17$). The results are summarized in (3), and portrayed graphically in Figure 3.

(3)   Mean ratings in the three conditions in word-likeness rating experiment

Condition	Token type	Rating
T~P	[stVt]	3.65
	*[spVp]	2.41
T~K	[stVt]	3.64
	*[skVk]	2.43
K~P	*[skVk]	2.52
	*[spVp]	2.41

Figure 3: Mean ratings for the three conditions in the English word-likeness rating experiment. Error bars show the 95% confidence intervals.
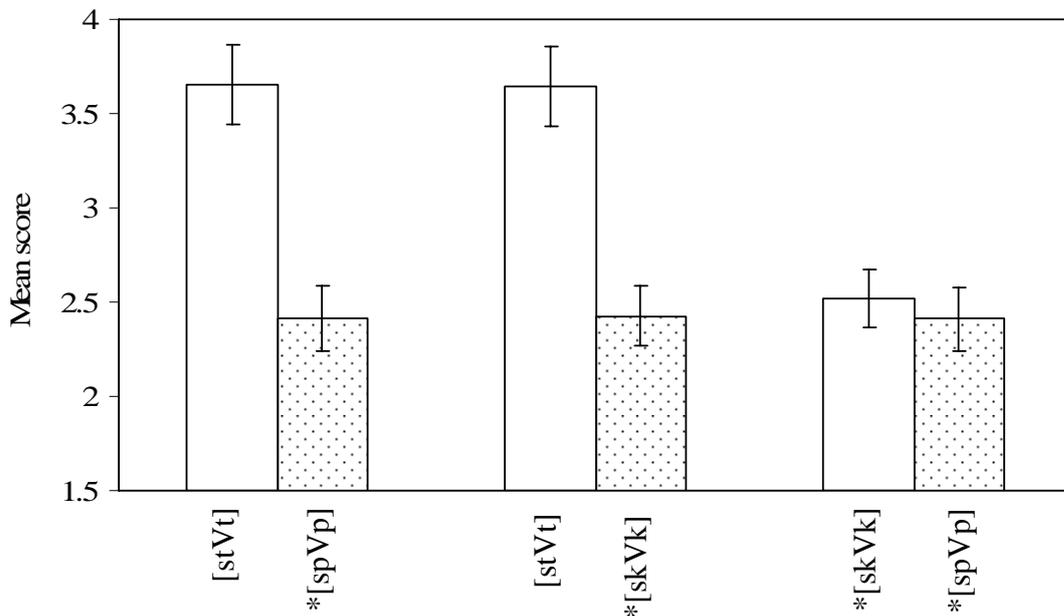
Like the Hebrew subjects, the English subjects rated grammatical forms better than ungrammatical forms. However, there is no significant difference in the scores of the two kinds of ungrammatical forms. This extends on Berent and Shimron's results. In their experiment, we saw that subjects do not distinguish between different kinds of grammatical forms in a word-likeness rating task. The results of this English experiment show the same for ungrammatical forms.

10

Comparative word-likeness

Design

Subjects. The same 20 subjects that took part in the word-likeness rating experiment also took part in the comparative word-likeness experiment.

Token selection. Tokens were selected in three conditions: (i) T~K: 15 non-word pairs of the form [stVt]~*[skVk]; (ii) T~P: 15 non-word pairs of the form [stVt]~*[spVp]; (iii) K~P: 15 non-word pairs of the form *[skVk]~*[spVp]. Tokens were selected such that lexical statistics (CBP and LND) and grammar conflicted in each non-word pair. In the T~K- and T~P-conditions, the LND and CBP of the [stVt]-token was lower than that of the *[skVk]- and *[spVp]-tokens respectively for each token-pair. In the K~P-condition, the lexical statistics of the *[skVk]-token was lower than that of the *[spVp]-token for each token-pair. The expected response pattern based on lexical statistics and that based on grammar are therefore directly opposite. The actual tokens used and their lexical statistics are included in the appendix.

Recordings. Recordings were done in exactly the same manner as for the word-likeness rating experiment.

Procedure. There were a total of 45 test-token pairs – 15 in each condition. I added 45 filler pairs to this. Only non-words were used in the filler pairs.[11] 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 who had to select the member of a pair that they thought to be most word-like. About 5 minutes elapsed between the presentation of the lists. On each presentation of a list, it was randomized differently. Before the list was presented the first time, 10 filler token-pairs were presented as practice trials. Since there is no correct or wrong answer, no feedback was given during the practice trials or during the actual experiment trials.

Results and discussion

Based on the results of Berent and Shimron, and the word-likeness rating experiment discussed above, we expect the subjects to prefer [stVt]-forms over *[skVk]- and *[spVp]-forms. We also expect subjects to make a distinction between *[skVk] and *[spVp] if there is indeed a well-formedness difference between these kinds of forms. As explained earlier, based on general patterns of consonant co-occurrence in English, we expect that *[skVk]-forms are more well-formed than *[spVp]-forms, even if both

11

of these are ungrammatical. We therefore expect that the subjects will prefer *[skVk] more often than *[spVp] when they have to choose between these two forms.

The results of the experiment were scored as follows: Each token pair, [Token 1]~[Token 2], was presented twice, so that there are three possible response patterns for each pair. If a subject selected [Token 1] more often than [Token 2], the [Token 1] was assigned a score of [1] for that subject, and [Token 2] was assigned a score of [0]. Conversely, if [Token 2] was selected more often than [Token 1], the [Token 2] received a score of [1] and [Token 1] received a score of [0]. If the tokens were selected with equal frequency, both were assigned a score of [1/2] for that subject. These scores were submitted to a 2×3 ANOVA with hypothesized grammatical well-formedness and condition as independent variables. A main effect of well-formedness was found ($F(1, 1794) = 68.5, p < 0.001$), as well as significant interaction between well-formedness and condition ($F(2, 1794) = 18.4, p < .001$). The contrasts between the more and less well-formed tokens in each condition were further investigated one-tailed paired sample t-tests. As before, I corrected for type 1 errors by dividing the critical p-value by the number of comparisons. All three comparisons returned significant results: T~P-condition ($t(299) = 15.4, p < .001$), T~K-condition ($t(299) = 11.9, p < .001$), and K~P-condition ($t(299) = 2.3, p = .01$). The results are summarized in (4), and portrayed graphically in Figure 4.

(4)   Percentage that token type was selected in comparative word-likeness experiment

Condition	Token type	Percentage
T~P	[stVt]	78
	*[spVp]	22
T~K	[stVt]	75
	*[skVk]	25
K~P	*[skVk]	55
	*[spVp]	45

Figure 4: Percentage that token type was selected in the English comparative word-likeness experiment. Error bars show 95% confidence intervals.
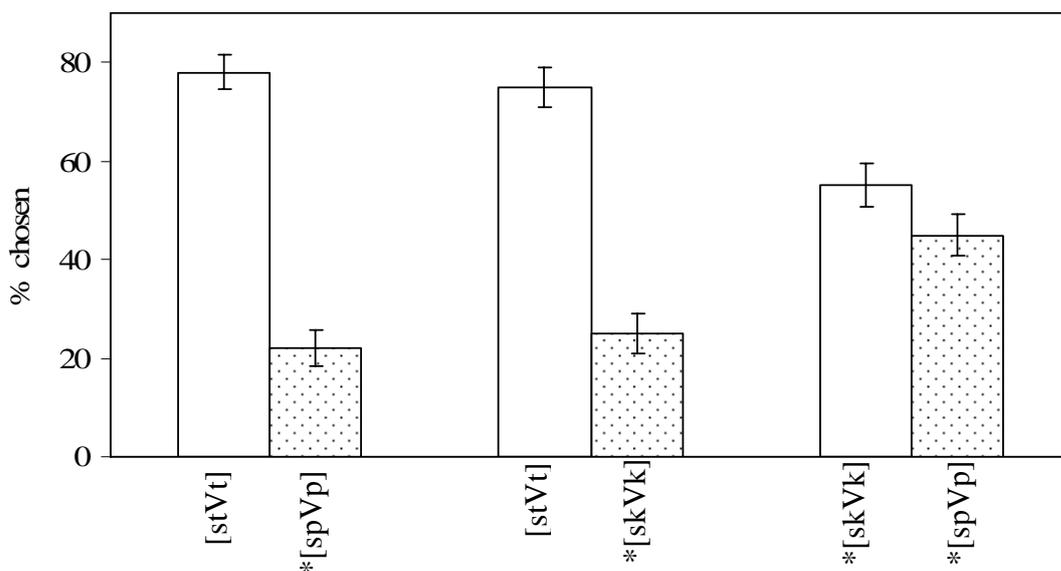
These results show that the grammatical [stVt]-forms were preferred over the two types of ungrammatical forms. This replicates the findings of Berent and Shimron for Hebrew and was also expected based on the results of the word-likeness rating experiment discussed above. However, the results of this experiment also extend on those of Berent and Shimron. The English listeners preferred the *[skVk]-forms over *[spVp]-forms, although neither of these are possible words of English. This shows that language users can make finer distinctions within the set of ungrammatical forms in terms of well-formedness.

Taken together, the results of Berent and Shimron on Hebrew and the results on English discussed here, show the following: language users can use the information provided by grammar in both a categorical and in a gradient manner. In task conditions that do not require explicit comparison between forms, language users make only the categorical distinction between grammatical and ungrammatical. However, if task conditions require an explicit comparison between forms, then language users make finer gradient distinctions within the sets of grammatical and ungrammatical forms. We therefore need a theory of grammar that can do both of these things. In the next section, I show how an OT grammar can be used to do just this.

An Optimality Theoretic account of categorical and gradient behavior

In this section, I will show that an OT grammar is ideally suited to model both kinds of behavior observed in the experiments discussed above. The following is what happens in a word-likeness rating task: when presented with a non-word, the language user asks the question: is there any possible input that my grammar will map onto this non-word token? An affirmative answer to this question means that the token is a possible word

and therefore grammatical. A negative answer means that it is not a possible word and therefore ungrammatical.

The evaluation proceeds differently in a comparative word-likeness task. Now the language user is presented with more than one non-word. The first step is to determine the input for each non-word that would result in the most harmonic mapping onto the non-word. (Crucially, it is not necessary that the mapping from the assumed input onto the non-word be possible in the language. All that is necessary is that it be the input that would result in the most harmonic mapping onto the non-word.) The language user then compares the input~output-mappings for each of the non-word tokens.

Hebrew

In Hebrew, non-geminates ([X-Y-Z]) and final-geminates ([X-Y-Y]) are grammatical, while initial-geminates (*[X-X-Y]) are ungrammatical. In this section, I first develop an OT account for this, and then show how this OT grammar can be used to explain the results of Berent and Shimron's experiments discussed above.[12]

Following McCarthy (1981, 1986), I assume a distinction between verbal roots and stems. The root is the bare consonantal form stored in the lexicon. The stem is the derived morphological category to which affixes attach, and consists of a combination of the root consonants and the vowels that express the specific conjugational class (*binyan*) of the verb.[13] Also following McCarthy, I will assume that final-geminate verbs are derived from bi-consonantal roots (i.e. /X-Y/ $\rightarrow$ [X-Y-Y] and not /X-Y-Y/ $\rightarrow$ [X-Y-Y]).[14] If these assumptions are made, then there are three questions that still need answering: (i) Why do bi-consonantal roots map onto tri-consonantal forms? Or: why does /X-Y/ not map onto *[X-Y]? (ii) Why do bi-consonantal roots only map onto final-geminates? Or: why does /X-Y/ not map onto *[X-X-Y]? (iii) Under richness of the base, we cannot exclude /X-Y-Y/ and /X-X-Y/ roots from the lexicon. If such roots do (or at least can) exist, then why do final and initial output geminates not originate from these forms? I discuss each of these questions in turn.

Let us first consider the question of why bi-consonantal roots map onto tri-consonantal stems. McCarthy and Prince (1990) argue that the verbal stem in Semitic languages must end on a consonant (see also Gafos, 1999, 2003). I follow them in assuming the existence a constraint FINAL-C.

(5)   **FINAL-C**: The verbal stem must end on a consonant.

This is the constraint that is responsible for forcing bi-consonantal roots to map onto tri-consonantal stems. To understand why, we have to consider the vowels that form part of the verbal stem. There are many conjugational classes in Hebrew that are expressed by a bi-vocalic melody. Consider the so-called *piʕel* as example. This conjugational class is characterized by the vocalic melody [i-e]. The stem of a Hebrew verb includes both the root consonants and the vocalic melody – that is, inflectional affixes are attached to the unit comprised of the root plus the vocalic melody. Without

14

augmenting a bi-consonantal root, the stem will end on a vowel – $/X^R\text{-}Y^R$, i-e/ then maps onto $*[|X^R i Y^R e|]$.[15, 16] FINAL-C therefore forces the augmentation of bi-consonantal roots by addition of a consonant. The actually observed mapping, $/X^R\text{-}Y^R$, i-e/ $\rightarrow$ $[|X^R i Y^R e Y|]$, of course violates the faithfulness constraint INTEGRITY (because input /Y/ has two surface correspondents).This gives evidence for the ranking FINAL-C >> INTEGRITY.

(6)   **INTEGRITY**:   Let S_I be the input and S_O the output such that $S_I \Re S_O$. No element of S_I has multiple correspondents in S_O.

For $x \in S_I$ and $w, z \in S_O$, if $x \Re w$ and $x \Re z$, then $w=z$ (McCarthy and Prince 1995).

Next consider the question of why bi-consonantal roots map onto final-geminates but never onto initial-geminates. Both $/X^R\text{-}Y^R/ \rightarrow [|X^R\text{-}Y_i^R\text{-}Y_i|]$[17] and $/X^R\text{-}Y^R/ \rightarrow *[|X_i\text{-}X_i^R\text{-}Y^R|]$ satisfy FINAL-C. Why then is the former grammatical but the latter ungrammatical? The difference between these structures is in the alignment of the root and the stem. In the grammatical $[|X^R\text{-}Y_i^R\text{-}Y_i|]$, the root and stem are perfectly aligned at their left edges. However, in the ungrammatical $*[|X_i\text{-}X_i^R\text{-}Y^R|]$, the root and stem are misaligned at their left edges. I argue that it is an alignment constraint that forces the final consonant to spread.

(7)   **ALIGN-L**:   The left edge of the root and the left edge of the stem must be aligned.

We now have all the rankings we need to explain why bi-consonantal roots map onto final-geminates. This is illustrated in the tableau in (8). Note that the ranking ALIGN-L relative to the other constraints does not matter.[18]

(8)   Bi-consonantal roots map onto final-geminates

$/X^R\text{-}Y^R/$		ALIGN-L	FINAL-C	INTEGRITY
a.	$|X^R\text{-}Y^R|$		*!	
b.	$|X_i\text{-}X_i^R\text{-}Y^R|$	*!		*
c. ☞	$|X^R\text{-}Y_i^R\text{-}Y_i|$			*

Finally, we need to explain why final-geminates and initial-geminates cannot originate from roots with final or initial identical consonants – i.e. why are both $/X^R\text{-}Y^R\text{-}Y^R/ \rightarrow$ $*[|X^R\text{-}Y^R\text{-}Y^R|]$ and $/X^R\text{-}X^R\text{-}Y^R/ \rightarrow *[|X^R\text{-}X^R\text{-}Y^R|]$ ungrammatical? Let us start with initial geminates. The ungrammatical form $*[|X^R\text{-}X^R\text{-}Y^R|]$ has two identical, contiguous consonants in the surface realization of the root. The avoidance of multiple occurrences of identical consonants in some domain can be explained with reference to the Obligatory Contour Principle (OCP) (McCarthy 1986). A version of the OCP indexed to the root is active in the phonology of Hebrew. This version of the OCP is defined in (9). Notice that this is a constraint on the surface realization of the root, and therefore

does not place a restriction on possible inputs. In the definition of this constraint, adjacency should be interpreted on the consonantal tier.

(9)   **OCP$_{\text{Root}}$**:   No contiguous, identical consonants in the surface realization of a root.

To explain the absence of surface forms like *[|X^R-X^R-Y^R|], OCP$_{\text{Root}}$ has to rank higher than some faithfulness constraint that can be violated in order to avoid violation of OCP$_{\text{Root}}$. Since roots like /X^R-X^R-Y^R/ do not actually exist, there is no evidence for what the relevant faithfulness constraint is. I will assume that one of the identical consonants in the root deletes, earning a violation of MAX-C, to avoid violation of OCP$_{\text{Root}}$. Of course, once one of the identical consonants has deleted, there are only two root consonants left. In order to satisfy the demands of FINAL-C, a third consonant needs to be supplied. We already know that FINAL-C outranks INTEGRITY, and consequently the extra consonant can be supplied by copying the final root consonant. The upshot is that an input like /X^R-X^R-Y^R/ will map onto the same output as a bi-consonantal input /X^R-Y^R/, i.e. [|X^R-Y$_i^R$-Y$_i$|]. This is shown in the tableau in (10). I do not include the candidate that copies the initial consonant and that is eliminated by ALIGN-L. Since geminate roots do not exist in the Hebrew lexicon, the inputs used in (10) are all hypothetical forms.

(10)   OCP$_{\text{Root}}$ >> {MAX-C, INTEGRITY}

			OCP$_{\text{Root}}$	FINAL-C	MAX-C	INTEGRITY
/X^R-X^R-Y^R/	a.	|X^R-X^R-Y^R|	*!			
	b.	|X^R-Y^R|		*!	*	
	c. ☞	|X^R-Y$_i^R$-Y$_i$|			*	*
/X^R-Y^R-Y^R/	d.	|X^R-Y^R-Y^R|	*!			
	e.	|X^R-Y^R|		*!	*	
	f. ☞	|X^R-Y$_i^R$-Y$_i$|			*	*

Of course, if [|X^R-X^R-Y^R|] violates OCP$_{\text{Root}}$, then so does [|X^R-Y^R-Y^R|]. An input root with identical final consonants will therefore be treated in exactly the same way as an input root with identical initial consonants – one of the consonants will delete, and in order to satisfy FINAL-C, the remaining final consonant doubles. This is also shown in the tableau in (10). Note that the faithful but ungrammatical candidate (10d) is phonetically identical to the optimal candidate (10f). The difference between these candidates lies in their hidden morphophonological structure.

There is considerable cross-linguistic evidence that consonantal co-occurrence constraints can apply to different morphological domains – see Tessier (2004) for a recent review. If OCP$_{\text{Root}}$ exists, then a similar constraint indexed to the larger morphological domain of the stem also exists – defined in (11). The last question to answer is where OCP$_{\text{Stem}}$ ranks in Hebrew. In order to avoid violation of FINAL-C,

16

Hebrew doubles the second consonant of a bi-consonantal root – see (8) above. In the observed output form, $[|X^R\text{-}Y_i^R\text{-}Y_i|]$, the sequence $[Y_i^R\text{-}Y_i]$ violates OCP$_{Stem}$. Hebrew therefore also tolerates violation of OCP$_{Stem}$ to avoid a FINAL-C violation, so that OCP$_{Stem}$ must also rank below FINAL-C. This is shown in the tableau in (12).

(11) **OCP$_{Stem}$:**   No contiguous, identical consonants in the surface realization of a stem.

(12)   FINAL-C >> OCP$_{Stem}$

/$X^R\text{-}Y^R$/	FINAL-C	INTEGRITY	OCP$_{Stem}$
$\mid X^R\text{-}Y^R \mid$	*!		
☞ $\mid X^R\text{-}Y_i^R\text{-}Y_i \mid$		*	*

We now have all the constraints and crucial rankings that we need to account for the distribution of contiguous identical consonants in Hebrew verbs. The tableau in (13) shows how this grammar will deal with the different possible inputs. Note that the ranking of ALIGN-L does not matter in this tableau – every candidate that violates ALIGN-L is harmonically bounded. Also note that because OCP$_{Root}$ and OCP$_{Stem}$ are in a stringency relationship, no ranking can be established between these two constraints with only the phenomena that we have considered above.

In a root with three non-identical consonants, the faithful candidate does not violate any constraints. Unsurprisingly, the faithful candidate (a) is therefore optimal. Now consider the bi-consonantal root input. Faithful (e) fatally violates FINAL-C. Both (f) and (g) satisfy FINAL-C by doubling one of the root consonants, earning them violations of INTEGRITY. However, (g) doubles the initial root consonant so that the root and stem are misaligned at their left edges in this candidate. This earns it a fatal violation of ALIGN-L. Finally, consider roots that contain contiguous identical consonants (either in initial or final position). The faithful candidates of both of these inputs, (h) and (l), fatally violate OCP$_{Root}$. This violation is avoided by deleting one of the identical consonants. The root is then treated just like a bi-consonantal root – in order to satisfy FINAL-C, the remaining final root consonant doubles.

17

(13)   Basic verbal grammar of Hebrew

Root structure			OCP_{Rt}	AL-L	FIN-C	OCP_{St}	INT	MAX-C
No identical: $/X^R\text{-}Y^R\text{-}Z^R/$	a.	☞ $\|X^R\text{-}Y^R\text{-}Z^R\|$						
	b.	$\|X^R\text{-}Y_i^R\text{-}Y_i\|$				*!	*!	*!
	c.	$\|X_i\text{-}X_i^R\text{-}Y^R\|$		*!		*	*	*
	d.	$\|X^R\text{-}Y^R\|$			*!			*
Bi-consonantal: $/X^R\text{-}Y^R/$	e.	$\|X^R\text{-}Y^R\|$			*!			
	f.	☞ $\|X^R\text{-}Y_i^R\text{-}Y_i\|$				*	*	
	g.	$\|X_i\text{-}X_i^R\text{-}Y^R\|$		*!		*	*	
Final identical: $/X^R\text{-}Y^R\text{-}Y^R/$	h.	$\|X^R\text{-}Y^R\text{-}Y^R\|$	*!			*		
	i.	$\|X^R\text{-}Y^R\|$			*!			*
	j.	☞ $\|X^R\text{-}Y_i^R\text{-}Y_i\|$				*	*	*
	k.	$\|X_i\text{-}X_i^R\text{-}Y^R\|$		*!		*	*	*
Initial identical: $/X^R\text{-}X^R\text{-}Y^R/$	l.	$\|X^R\text{-}X^R\text{-}Y^R\|$	*!			*		
	m.	$\|X^R\text{-}Y^R\|$			*!			*
	n.	☞ $\|X^R\text{-}Y_i^R\text{-}Y_i\|$				*	*	*
	o.	$\|X_i\text{-}X_i^R\text{-}Y^R\|$		*!		*	*	*

Now that we have a grammar for Hebrew, we can look at the results of Berent and Shimron again. Consider first the results of their word-likeness rating experiment. In this experiment, they found that Hebrew speakers rated the two kinds of grammatical non-words ([X-Y-Z] and [X-Y-Y]) better than the ungrammatical non-words (*[X-X-Y]), but they did not distinguish between the different grammatical non-words. We can explain this as follows: when the subjects are required to rate a single non-word at a time, they determine whether there is at least one input that their grammar can map grammatically onto that non-word. This is equivalent to asking whether the non-word is a possible word or not. If there is an input that would map grammatically onto the non-word, the subjects assign it a high score. On the other hand, if there is no such input, subjects assign the non-word a low score.

As the tableau in (13) shows, there is some input that will map onto a non-geminate – namely an input identical to the output. /X-Y-Z/ → [X-Y-Z] is grammatical, and therefore non-geminate non-words are identified as possible words and assigned high scores. Tableau (13) also shows that there are inputs that will be mapped grammatically onto final-geminates. In fact, there are three such inputs, namely /X-Y/, /X-Y-Y/ and /X-X-Y/. Also final-geminates are then identified as possible words and assigned high ratings. The situation with initial-geminates is different. None of the logically possible inputs will be mapped grammatically onto a form such as *[X-X-Y]. This is because

both /X-X-Y/ and /X-Y/ are mapped onto final-geminates. Initial-geminates are then identified as impossible words, and assigned low scores.

Now consider the comparative word-likeness experiment. In this experiment, Berent and Shimron found that their subjects imposed the following word-likeness hierarchy on the non-word tokens: non-geminates > final-geminates > initial-geminates. I propose that language users do this as follows: for each of the non-words, they find the input that would most harmonically map onto the non-word. We can identify these inputs in the tableau in (13). For the non-geminate [|X-Y-Z|], this input is obviously /X^R-Y^R-Z^R/. The input that results in the most harmonic mapping onto a final-geminate is the bi-consonantal root – the mapping /X^R-Y^R/ $\rightarrow$ [|X^R-Y_i^R-Y_i|] violates INTEGRITY and OCP$_{Stem}$. All other mappings onto a final-geminate violate either a superset of these two constraints or violate high ranking OCP$_{Root}$. The input that most harmonically maps onto an initial-geminate is either /X^R-Y^R/ or /X^R-X^R-Y^R/ – the choice depends on the ranking between ALIGN-L and OCP$_{Root}$, and since neither of these two are violated in Hebrew we cannot determine the ranking between them. For the sake of simplicity, I consider only /X^R-X^R-Y^R/ in the rest of the discussion. However, the same result would be achieved if /X^R-Y^R/ were used.

Once the input that results in the most harmonic mapping for each non-word has been determined, the language user compares the mappings in a "comparative tableau", as shown in (14).

(14)   Comparing the different non-words in Hebrew

			OCP$_{Rt}$	AL-L	FIN-C	OCP$_{St}$	INT	MAX-C
NoG	1	/X-Y-Z/ $\rightarrow$ [|X^R-Y^R-Z^R|]						
FinG	2	/X-Y/ $\rightarrow$ [|X^R-Y_i^R-Y_i|]				*	*	
IniG	3	/X-X-Y/ $\rightarrow$ [|X^R-X^R-Y^R|]	*			*		

The comparative tableau in (14) is not an ordinary OT tableau – it does not compare different output candidates for the same input, but rather different input~output-mappings.[19] Rather than the usual pointing hand to indicate the winning candidate, I use Arabic numerals to indicate the order that the grammar imposes on these forms. The non-geminate mapping violates none of the constraints, while the final-geminate mapping violates OCP$_{Stem}$ and INTEGRITY. When these two mappings are compared, the non-geminate mapping is therefore the more well-formed option. Although both of these represent mappings that are possible in Hebrew, they are not equally well-formed. The non-geminate form is perfectly unmarked and perfectly faithful, while the final-geminate is neither perfectly faithful nor unmarked. This corresponds to the way in which Berent and Shimron's subjects responded. When required to compare non-geminates and final-geminates, they preferred the non-geminates over the final-geminates.

19

The initial-geminate form violates OCP$_{Root}$. Because of the ranking OCP$_{Root}$ >> INTEGRITY, final-geminates are better than initial-geminates, and this is again how they were rated by the subjects in the experiment.

A note is in order here about how to interpret the tableau in (14). This tableau shows that non-geminates are more well-formed than final-geminates which are again more well-formed than initial-geminates. I claim that the subjects used this information as follows: when asked to choose between a non-geminate and a final-geminate, the subjects are *more likely* to select the non-geminate. In addition to grammar, there are other factors that influence how subjects respond. These other factors include things such as the lexical statistics of tokens (see the earlier discussion), but also things like fatigue, lack of concentration, and individual differences. We can therefore not expect that subjects will *always* act according the information provided by grammar. However, we can expect that the information provided by grammar will bias their responses.

Since OT already has the ability to compare forms for their relative well-formedness, it is a small change that has to be made to the theory to explain what the subjects do in a comparative word-likeness experiment. Rather than comparing output candidates for the same input, they compare output candidates that do not share the same input.

English

English allows words of the form [stVt], but not of the form *[skVk] or *[spVp]. In this section, I first develop an OT account to explain this, and then show how this account explains the results of the word-likeness experiments that I discussed earlier.

I assume a markedness constraint against each of the three kinds of tokens considered here, i.e. *stVt, *skVk, and *spVp. These constraints can be viewed as the local conjunction of the OCP-type constraints against multiple occurrences [t], [k] or [p] in a single syllable, with a constraint against the sequence [s+stop] (Alderete, 1997; Coetzee, 2004:402-420; Pater and Coetzee, 2005).[20]

It is well established that languages often place restrictions on the occurrence of multiple identical consonants in certain local domains. The Hebrew data discussed just above is an example of this. Similar restrictions also hold in languages as diverse as Arabic (Frisch *et al*., 2004; McCarthy, 1994), Japanese (Kawahara *et al*., 2006), Russian (Padgett, 1995), Muna (Coetzee and Pater, 2006), French, Latin and English (Berkley, 2000). In all of these languages, there is evidence that identical (or highly similar) consonants are avoided in some local context. Although there are differences in the details of these restrictions in the different languages, it is clear that there must be a general family of constraints that militate against multiple occurrences of identical (or similar) consonants. I propose that the versions of these constraints stated in (15) are relevant to the restriction in English that is our focus here.[21]

20

(15) *[t…t]$_\sigma$: Do not allow a syllable with two [t]'s.
      *[k…k]$_\sigma$: Do not allow a syllable with two [k]'s.
      *[p…p]$_\sigma$: Do not allow a syllable with two [p]'s.

There is an extensive literature on the markedness of [s+stop] sequences (Broselow, 1991; Davis, 1984; Hayes, 1985:140-149; Kahn, 1980; Lamontagne, 1993: Chapter 6; Morelli, 1999; Selkirk, 1982; etc.). There are several views on why this sequence is marked. One view considers these structures to be true consonant clusters. If this is the case, then they violate the Sonority Sequencing Principle – sonority falls from [s] to the [stop]. Under another view, these structures are complex segments rather than consonant clusters. This avoids the violation of the SSP. However, if they are complex segments then they are marked per se. Either way, an [s+stop]-sequence is marked. In (16), I formulate a constraint against such sequences.

(16) ***[s+stop]$_\sigma$**: [s] is not allowed to immediately precede a stop in one syllable.

English violates each of the constraints in (15) and (16) individually – this is evidenced by the existence of words such as *sty* (*[s+stop]$_\sigma$), *toot* (*[t…t]$_\sigma$), *cake* (*[k…k]$_\sigma$), and *pop* (*[p…p]$_\sigma$). This means that all of these constraints rank very low in English – so low that they have no discernable effect. What English does not tolerate is the violation of certain combinations of these constraints within a single syllable – specifically, violation of *[s+stop]$_\sigma$ and any of *[k…k]$_\sigma$ or *[p…p]$_\sigma$. OT accounts for the avoidance of the local accumulation of marked structure through local conjunction of markedness constraints (Smolensky, 1995). We can now show that the *sCVC constraints are just the local conjunction of each of the constraints in (15) with the constraint in (16).

(17) *stVt: Do not violate *[s+stop]$_\sigma$ and *[t…t]$_\sigma$ in the same syllable.
      *skVk: Do not violate *[s+stop]$_\sigma$ and *[k…k]$_\sigma$ in the same syllable.
      *spVp: Do not violate *[s+stop]$_\sigma$ and *[p…p]$_\sigma$ in the same syllable.

We know that English tolerates violation of *stVt but not of *skVk or *spVp. This gives evidence for the ranking {*spVp, *skVk} >> Faithfulness >> *stVt. But what about the ranking between *spVp and *skVk? As I have shown above, there are more words in English with two dorsals in one syllable than words with two labials in one syllable. This can be captured by the ranking *[p…p]$_\sigma$ >> *[k…k]$_\sigma$. A question that needs to be answered is how this ranking can be learned. Since English tolerates violation of both these constraints, classical error-driven learning (Tesar and Smolensky, 1998) cannot be used. Pater (2005) develops an algorithm that can learn rankings from statistical patterns in the lexicon, and that is hence ideally suited for this kind of scenario. When two markedness constraints that are both freely violated in some language have to be ranked, Pater's algorithm ranks higher the constraint that is violated less often. When his algorithm has to rank *[p…p]$_\sigma$ and *[k…k]$_\sigma$, it will hence rank *[p…p]$_\sigma$ higher, since this is the constraint that is violated less often in the English lexicon. Once the ranking *[p…p]$_\sigma$ >> *[k…k]$_\sigma$ has been learned, Itô and Mester's (2003) principle of "ranking preservation" can be used to infer the ranking *spVp >> *skVk. This principle requires the following: let *LC₁* and *LC₂* 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 $C_1 \gg C_2$, then $LC_1 \gg LC_2$. Given that *[p…p]$_\sigma$ and *[k…k]$_\sigma$ are conjuncts of *spVp and *skVk respectively, and given the ranking *[p…p]$_\sigma$ $\gg$ *[k…k]$_\sigma$, it follows by ranking preservation that *spVp $\gg$ *skVk.

We now have the following mini-grammar for English: *spVp $\gg$ *skVk $\gg$ Faithfulness $\gg$ *stVt. The tableau in (18) shows that this grammar does correctly predict that [stVt] is a possible word of English while neither *[skVk] nor *[spVp] is. In this tableau, I assume that the relevant faithfulness constraint is IDENT[place]. However, since there are no active alternations in English involving *[spVp] and *[skVk], we cannot know for sure what the constraint should be. It can be any faithfulness constraint that can be violated to avoid violation of *[skVk] and *[spVp].

(18)   Mini-grammar of English [sCVC]-forms

		*spVp	*skVk	IDENT[place]	*stVt
/stVt/	☞ stVt				*
	stVk			*!	
/skVk/	skVk		*!		
	☞ skVt			*	
/spVp/	spVp	*!			
	☞ spVt			*	

This tableau shows that an /stVt/-input will map faithfully onto itself, and therefore [stVt] is correctly predicted as a possible word. However, neither /skVk/ nor /spVp/ maps faithfully onto itself. *[skVk] and *[spVp] are therefore correctly predicted not to be possible words. We now have all the information we need to explain the results of the English experiments discussed above.

Consider first the word-likeness rating experiment. In this experiment, subjects had to rate non-words individually for their word-likeness – i.e. no direct comparison between non-words was required. The results show that the subjects made a categorical distinction between possible and impossible words – the grammatical [stVt] non-words received high scores, and the ungrammatical *[skVk] and *[spVp] forms received low scores but were not distinguished from each other. The subjects used their grammar in the same manner as the subjects in Berent and Shimron's experiment. For each non-word, they determine whether there is an input that will map unto the specific non-word. If such an input exists, the non-word is identified as a possible word, and it receives a good rating. Non-words for which no such input exists are identified as impossible words, and receive low scores.

Now consider the comparative word-likeness experiment. In this experiment, the subjects were required to compare two non-words and to select the one that they deemed most word-like. The results of the experiment show that the subjects rated the

non-words as follows: [stVt] > *[skVk] > *[spVp]. They no longer divide the set of non-words into the two categorical classes of grammatical and ungrammatical. Rather, the non-words are now rated gradiently according to their relative well-formedness. The subjects again responded like the subjects in Berent and Shimron's experiment. For each non-word, they first determine the input that would result in the most harmonic mapping onto the non-word. For an [stVt] non-word, this input is obviously /stVt/. Any other input that maps onto [stVt] will violate some faithfulness constraint in addition to *stVt. Similarly, the relevant input for *[skVk] will be /skVk/ – again because any other input will violate a faithfulness constraint in addition to *skVk. With similar reasoning, we can also show that the relevant input for *[spVp] will be /spVp/. Once these inputs have been determined, the subjects compare the three input~output-mappings in a comparative tableau. This comparison is shown in (19).

(19)   Comparing non-words in English

		*spVp	*skVk	Faithfulness	*stVt
1	/stVt/ → [stVt]				*
2	/skVk/ → [skVk]		*		
3	/spVp/ → [spVp]	*			

As before, this comparative tableau should be interpreted differently from a standard OT tableau. This is not a production oriented tableau in which different output candidates for the same input are compared. We are not interested in the best candidate, but rather in how all three candidates are related to each other. Since /stVt/ → [stVt] violates the lowest ranking constraint, it is rated best, as indicated by the numeral [1] next to this mapping. Since /spVp/ → [spVp] violates the highest ranking constraint, it is rated worst of all, indicated by the [3] next to this mapping. This corresponds to how the subjects rated these tokens in the comparative word-likeness experiment.

Considering alternatives

In the previous sections, I have shown that an OT grammar can account for both the categorical and the gradient response patterns observed in the experiments. In this brief section, I will show that standard generative grammars have no problems in accounting for the categorical grammatical/ungrammatical distinction, but that they are less well-suited to make the finer gradient distinctions between degrees of grammaticality and ungrammaticality. I will also show that models in which well-formedness ratings are a direct reflex of gradient lexical statistics cannot account for the response patterns.

Standard generative models of grammar have no difficulty in differentiating grammatical forms from ungrammatical forms. An ungrammatical form is a form that cannot be generated as grammatical output from any permissible input, and all other forms are grammatical. One reason why some form cannot be generated as a grammatical output is that the input required to derive it is absent from the lexicon because it violates a morpheme structure constraint (MSC). Davis (1984), for instance,

23

claims that *[skVk] and *[spVp] are ungrammatical in English because there is an MSC that bans /skVk/ and /spVp/ from the English lexicon. He defines this MSC as in (20) (Davis, 1984:46). This MSC ensures that [stVt] is grammatical while *[skVk] and *[spVp] are not.

(20)

$$\ast s \begin{bmatrix} \text{-cont} \\ \text{-cor} \\ \alpha\text{voice} \\ \beta\text{ant} \end{bmatrix} \quad V \quad \begin{bmatrix} \text{-cont} \\ \text{-cor} \\ \alpha\text{voice} \\ \beta\text{ant} \end{bmatrix}$$

While (20) can distinguish between grammatical [stVt] and ungrammatical *[skVk] and *[spVp], it cannot distinguish between *[skVk] and *[spVp] in terms of their relative well-formedness. It is possible to split (20) into separate MSC's for labials and dorsals. *[skVk]-forms and *[spVp]-forms could then be distinguished from each other because they would be ruled out by different MSC's. However, we would still not be able to compare them for their relative well-formedness. To achieve that, we will also have to add comparative powers to the grammar – specifically, we have to impose an ordering between the dorsal-MSC and the labial-MSC such that violating the latter is deemed worse than violating the former. Although a standard generative grammar can in principle derive the well-formedness difference between *[spVp] and *[skVk], it needs to be embellished with special comparative powers to do so. Because OT is by design a comparative theory of grammar based on ranked constraints, the ability to distinguish between *[spVp] and *[skVk] in terms of relative well-formedness follows directly from the basic architecture of the grammar. No additional embellishments are necessary.

Another possible explanation for the response patterns observed in the experiments should be considered, namely the possibility that the results originated not in grammar but rather in statistical patterns from the lexicon. Hay *et al.* (2004), for instance, argue that "phonological grammar is a simple projection of the lexical statistics" (p. 59), and moreover that it "is gradient rather than categorical" (p. 71). The results of the Hebrew and English experiments discussed above support neither of these claims. First, it is clear that the results do not support the claim that well-formedness is (only) gradient. Berent and Shimron found in their experiments that exactly the same tokens are sometimes treated categorically the same, and sometimes gradiently distinguished from each other. The English experiments confirmed this.

Secondly, there is strong evidence that the results of the two sets of experiments cannot be explained solely from the lexical statistics. Consider the Hebrew experiments first. Again, since the same tokens were used in both experiments, the lexical statistics of the tokens in the experiments did not differ. If lexical statistics were responsible for the response patterns, then the subjects should have responded the same in both tasks.

In the design of the English experiments, I controlled specifically for the potential influence of lexical statistics. The tokens were all selected such that the lexical statistics between token types in a condition did not differ, or such that the lexical statistics conflicted with grammar (i.e. if grammar favored *x* over *y*, then the lexical statistics
24

favored *y* over *x*). If it were lexical statistics that determined the response patterns, then the subjects should either not have distinguished between the token types at all, or if they did make a distinction it should have been the opposite of what would be expected based on grammar alone. The results of the experiments clearly showed that the subjects did distinguish between the token types in well-formedness, and moreover the distinctions they made agreed with grammar and not with the lexical statistics. The response patterns therefore did not originate in the lexical statistics. Regression analyses confirm this. I performed a regression on the response patterns using LND and CBP as the independent variable. The results of these analyses for the two experiments are shown in (21). It is clear that the lexical statistics do not account for a significant part of the variation in the response patterns.

(21)   Regression analyses

	LND	CBP
Word-likeness rating	$r^2 = .01$	$r^2 < .01$
Comparative word-likeness	$r^2 = .02$	$r^2 < .01$

The results of the English experiment therefore differ from many results reported in the literature where word-likeness ratings did correspond more significantly with lexical statistics (Bailey and Hahn, 1998; Hay *et al.*, 2004; Coleman and Pierrehumbert, 1997). I cannot offer a clear reason for this difference in the results. However, the results reported in this paper show that the relationship between phonological grammar and the lexicon is more complicated than what is assumed by, for instance, Hay *et al.* (2004). More research is necessary before we can decide that phonological grammar is only a gradient projection from the lexicon.

Conclusion

In this paper, I have discussed data from two sets of word-likeness experiments, showing that humans use their grammar in both categorical and gradient ways. They do have strong intuitions about whether some non-word is a possible word or not (i.e. whether a form is grammatical or not). In task conditions that do not require explicit comparison between forms, the subjects in word-likeness experiments correspond according to these intuitions. This shows that the information provided by grammar can be interpreted in terms of the categorical distinction between grammatical and ungrammatical. But in addition to this distinction, humans can also make more fine grained gradient distinctions between more (im)possible and less (im)possible. They can compare two grammatical forms, and decide which is more word-like. Similarly, they can compare two ungrammatical forms and decide which is more word-like. In task conditions that require subjects to make these kinds of comparisons, they respond along these lines.

If language users can use the information provided by grammar to respond both in categorical and in gradient manners, then grammar should be able to provide both categorical and gradient information. This poses a challenge to standard generative

25

grammar, which was not designed as a comparative theory of grammar. However, because of its inherent comparative nature, OT is perfectly suited for this task. We can make the categorical distinction between grammatical and ungrammatical non-words as follows in an OT grammar: for a non-word, if there is an input that would be mapped grammatically onto the non-word, then the non-word is grammatical (is a possible word of the language). On the other hand, if there is no input that would map onto the non-word, then it is ungrammatical (not a possible word of the language). This is equivalent to what standard generative grammar does – asking whether there is a well-formed derivational history for the non-word. In OT, this is easily done using OT tableaux in the standard manner.

Since an OT grammar is by design comparative, it is also a straightforward matter to get information about the finer gradient well-formedness relationship between non-words. This can be done as follows: for every non-word, determine that input that will map most harmonically onto the specific non-word. Then compare the input~output-mapping for the different non-words in a comparative tableau. In this kind of tableau, EVAL does not compare different output candidates for a single input, but rather input~output-mappings that differ in terms of both input and output.

The results of the experiments also give evidence that phonological grammar is more than a mere projection of lexical statistics. The subjects in the experiments responded in ways that do not support a model where word-likeness is a direct mapping from lexical statistics. Although lexical statistics are gradient and continuous, subjects sometimes responded categorically, grouping together tokens that differ in terms of their lexical statistics. Sometimes the subjects also responded in ways directly opposite to what would have been expected based on the lexical statistics.

Having established that processing data reflect the influence of grammar, and having shown that we can account for this influence in OT, we can now use processing data as a rich and largely untapped source of information about grammatical competence. This opens up many new and interesting research possibilities.

Notes

26

2.   It is worth remarking that I use the term "gradient" here differently than, for instance, Boersma (1998) and Flemming (2001) do. My use of the term does not refer to gradient phonetic effects in the realization of categorical phonological categories, and it does not refer to different possible realizations of the same input. I use the term here to refer to gradient well-formedness distinctions that hold between forms that do not share the same input.

3.   I use X, Y and Z as variables that range over all of the Hebrew consonants. Real Hebrew words, of course, also contain vowels. Since the vowels are not relevant to the point made here, I abstract away from the vowels.

4.   I discuss only the most relevant aspects of the experimental design. Refer to Berent and Shimron (1997) for the full details. For discussion of the lexical statistics of the tokens, see more below.

5.   These difference scores were computed as follows (here and in the rest of the paper NoG = no geminate, IniG= initial geminate, and FinG= final geminate) :
(i) Difference Score (IniG~FinG) = Mean Score (FinG) – Mean Score (IniG).
(ii) Difference Score (IniG~NoG) = Mean Score (NoG) – Mean Score (IniG).
(iii) Difference Score (FinG~NoG) = Mean Score (NoG) – Mean Score (FinG).

6.   Berent and Shimron 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.

7.   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.

8.   The lexical statistics were all calculated from the CELEX database (Baayen *et al*., 1995). Since the phonetic transcriptions in CELEX reflect British pronunciation, this database was "Americanized" before the calculations were done. The changes that were made include things such as replacement of [a] with [æ] in words like *half*, addition of [ɹ] in the pronunciation of words like *car*, etc. I am indebted to John Kingston for this.

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

To understand how transitional probabilities were calculated, consider the token [skɔk] as an example. For the sequence [sk] we can calculate the probability of an [s] being followed by a [k], and the probability of a [k] being preceded by an

[s]. To calculate the probability of [s] being followed by [k]: (i) take the log of the frequency of [s]; (ii) take the log of the frequency of 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. The CBP of some token is the product of all the bi-phone probabilities of that token.

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

10.   The fillers were selected such that approximately an equal number of all tokens were possible words and impossible words. Fillers that represented impossible words violated a constraint on the consonants that co-occur in the onset and coda of a single syllable (Fudge, 1969). They were therefore ill-formed for reasons similar to the ill-formedness of the *[skVk] and *[spVp]-tokens.

11.   The fillers in this experiment were selected using the same criteria as in the previous experiment. See previous footnote.

12.   I sketch only the outlines of an account here. See Coetzee (2004:348--80) for more detail.

13.   The question of whether the root exists as a separate morphological entity has been questioned in recent years. Bat-El (1994), Ussishkin (1999) and Gafos (2003) do not assume the existence of the root. However, see Berent, Everett and Shimron (2001) and Berent, Shimron and Vaknin (2001) for arguments in favor of the root.

14.   This assumption is shared in by *inter alia* Gafos (1999). Ussishkin (1999) also assumes that geminate verbs derive from bi-consonantal forms. However, for him the bi-consonantal form is not a bare consonantal root but rather an output base that contains consonants and vowels. Gafos (2003) also assumes that geminates derive from bi-consonantal inputs. However, he assumes that the consonant that shows up as geminate on the surface is underlyingly linked to two moras.

15.   Stem boundaries are marked by vertical lines |. Superscripted R indicates morphological membership of the root.

16.   Other alternatives are ruled out by other high ranking constraints. $[|i.X^ReY^R|]$ and $[|X^Ri.eY^R|]$ are both ruled out by ONSET. $[|X^ReY^R|]$ and $[|X^RiY^R|]$ are ruled out by a constraint requiring faithful parsing of the vocalic melody – some version of MAX-V (Gafos, 2003; Ussishkin, 1999).

17.   Subscripted *i* represents phonological relatedness.

28

18.   There are several other candidates satisfying FINAL-C and ALIGN-L that are ruled out by high ranking constraints not considered here. First, there is a candidate that supplies the third stem-final consonant by epenthesis rather than copying, i.e. *[|X^R-Y^R-Z|]. The fact that Hebrew chooses copying over epenthesis, shows that DEP-C >> INTEGRITY. There is also a candidate that copies the initial root consonant to the right and a candidate that copies the final consonant to the left, i.e. [X_i^R-X_i-Y^R] and [X^R-Y_i-Y_i^R]. These candidates are ruled out by a high ranking constraint requiring the surface correspondents of the root to be contiguous, i.e. CONTIGUITY indexed to the root.

19.   See Berent and Shimron (1997) for a similar proposal, and Coetzee (2004) for an explicit formalization of an OT model that can do these kinds of comparisons. Sorace and Keller (2005:1516) claim that an OT grammar cannot compare candidates that are derived from different inputs. It is true that this kind of comparison is not usually done in production oriented OT. However, it is untrue that an OT grammar cannot do this. Nothing in the way that EVAL works depends on the origin of the candidates being compared – i.e. it can compare forms that do not share the same input. Even in classic OT, there is acknowledgement that EVAL can do this. In "lexicon optimization", EVAL compares different possible inputs for a single output (Prince and Smolensky, 1993). The tableau des tableaux introduced by Itô, Mester and Padgett (1995) gives formal expression of this property of EVAL in classical OT. See Coetzee (2004: Chapter 2) for a detailed discussion of this characteristic of EVAL.

20.   I give only a very basic motivation for these constraints here. See Coetzee (2004) for a complete discussion and motivation. See Baertsch and Davis (2003) for a very different approach.

21.   These constraints are too specific. The domain should probably be defined more broadly – as English also does not allow words of the form *[spV.pV] or *[skV.kV]. These constraints should also apply to more than just the voiceless stops – since English also does not tolerate *[slVl], *[snVn], etc. For more on these kinds of restrictions, see Fudge (1969). For discussion about how to state the domain of these constraints, see Coetzee (2004:420).

Appendix

Tokens used in English word-likeness rating experiment

T~K-condition

[sTvT]	LND	CBP	[sKvK]	LND	CBP
stʌt	10.65	0.066	skeɪk	26.06	0.186
stɑ:t	44.16	0.279	skæk	22.12	0.227
stɔ:t	23.06	0.217	skɑ:k	18.41	0.219
stʊt	17.06	0.144	skɛk	12.65	0.193
stʌt	42.19	0.219	skɪk	37.31	0.299
Mean	27.424	0.185		23.31	0.225

t-test on LND: *t*(8) = .52, *p* =.62   *t*-test on CBP: *t*(8) = .95, *p* =.37

T~P-condition

[sTvT]	LND	CBP	[sPvP]	LND	CBP
stʊt	17.06	0.144	spɑ:p	21.41	0.244
stɔɪt	10.65	0.066	spɛp	20.32	0.208
stʌt	42.19	0.219	spɪp	26.69	0.212
stɑ:t	44.16	0.279	spæp	17.98	0.201
stɔ:t	23.06	0.217	spi:p	27.45	0.213
	27.42	0.185		22.77	0.216

t-test on LND: *t*(8) = .67, *p* =.52   *t*-test on CBP: *t*(8) = .82, *p* =.44

K~P-condition

[sKvK]	LND	CBP	[sPvP]	LND	CBP
skaɪk	26.06	0.185	spæp	17.98	0.201
skaʊk	9.95	0.001	spɑ:p	21.41	0.244
ski:k	28.63	0.134	spɛp	20.32	0.208
skɔ:k	10.04	0.171	spɪp	26.69	0.212
sku:k	12.64	0.193	spi:p	27.45	0.213
	17.46	0.137		22.77	0.216

t-test on LND: *t*(8) = 1.18, *p* =.27   *t*-test on CBP: *t*(8) = 2.18, *p* =.06

Tokens used in English comparative word-likeness experiment

T~K-condition

[sTvT]	LND	CBP	[sKvK]	LND	CBP	Diff in LND	Diff in CBP
stɔːt	23.06	0.217	skæk	22.12	0.227	0.94	-0.010
stʊt	17.06	0.144	skɑːk	18.41	0.219	-1.35	-0.075
stɔɪt	10.65	0.066	skɛk	12.65	0.193	-2.00	-0.127
stɔɪt	10.65	0.066	skɑːk	18.41	0.219	-7.76	-0.153
stɔɪt	10.65	0.066	skʌk	19.38	0.183	-8.73	-0.117
stɔɪt	10.65	0.066	skæk	22.12	0.227	-11.47	-0.161
stʊt	17.06	0.144	skeɪk	26.06	0.186	-9.00	-0.042
stʌt	42.19	0.219	skɪk	37.31	0.299	4.88	-0.080
stɔːt	23.06	0.217	skɪk	37.31	0.299	-14.25	-0.082
stʊt	17.06	0.144	skɪk	37.31	0.299	-20.25	-0.155
stʊt	17.06	0.144	skæk	22.12	0.227	-5.06	-0.083
stɔɪt	10.65	0.066	skaɪk	13.89	0.157	-3.24	-0.091
stɔɪt	10.65	0.066	skeɪk	26.06	0.186	-15.41	-0.120
stʊt	17.06	0.144	skɛk	12.65	0.193	4.41	-0.049
stɔɪt	10.65	0.066	skɪk	37.31	0.299	-26.66	-0.233
	16.54	0.122		24.21	0.228		

t-test on LND: *t*(14) = 3.32, *p* =.005   *t*-test on CBP: *t*(14) = 7.26, *p* <.001

T~P-condition

[sTvT]	LND	CBP	[sPvP]	LND	CBP	Diff in LND	Diff in CBP
stɔːt	23.06	0.217	spɪp	26.69	0.212	-3.63	0.005
stɔɪt	10.65	0.066	spʌp	11.38	0.182	-0.73	-0.116
stɔɪt	10.65	0.066	spiːp	27.45	0.213	-16.80	-0.147
stɔɪt	10.65	0.066	spɪp	26.69	0.212	-16.04	-0.146
stɔːt	23.06	0.217	spiːp	27.45	0.213	-4.39	0.004
stɔɪt	10.65	0.066	spɑːp	21.41	0.244	-10.76	-0.178
stɔɪt	10.65	0.066	spæp	17.98	0.201	-7.33	-0.135
stʊt	17.06	0.144	spæp	17.98	0.201	-0.92	-0.057
stʊt	17.06	0.144	spɪp	26.69	0.212	-9.63	-0.068
stɔɪt	10.65	0.066	speɪp	17.13	0.163	-6.48	-0.097
stɔɪt	10.65	0.066	spɛp	20.32	0.208	-9.67	-0.142
stʊt	17.06	0.144	spɑːp	21.41	0.244	-4.35	-0.100
stʊt	17.06	0.144	spiːp	27.45	0.213	-10.39	-0.069
stʊt	17.06	0.144	spɛp	20.32	0.208	-3.26	-0.064
stʊt	17.06	0.144	spʌp	11.38	0.182	5.68	-0.038
	14.87	0.117		21.45	0.207		

t-test on LND: *t*(14) = 4.30, *p* <.001   *t*-test on CBP: *t*(14) = 6.28, *p* <.001

31

K~P-condition

[sKvK]	LND	CBP	[sPvP]	LND	CBP	Diff in LND	Diff in CBP
skaʊk	9.95	0.001	spæp	17.98	0.201	-8.03	-0.200
sku:k	12.64	0.193	spæp	17.98	0.201	-5.34	-0.008
skaʊk	9.95	0.001	spɑ:p	21.41	0.244	-11.46	-0.243
sku:k	12.64	0.193	spɑ:p	21.41	0.244	-8.77	-0.051
skʊk	3.86	0.097	spɑ:p	21.41	0.244	-17.55	-0.147
skaʊk	9.95	0.001	spɛp	20.32	0.208	-10.37	-0.207
sku:k	12.64	0.193	spɛp	20.32	0.208	-7.68	-0.015
skʊk	3.86	0.097	spɛp	20.32	0.208	-16.46	-0.111
ski:k	28.63	0.134	spɪp	26.69	0.212	1.94	-0.078
skɔ:k	10.04	0.171	spɪp	26.69	0.212	-16.65	-0.041
skʊk	3.86	0.097	spɪp	26.69	0.212	-22.83	-0.115
skɔ:k	10.04	0.171	spi:p	27.45	0.213	-17.41	-0.042
skʊk	3.86	0.097	spi:p	27.45	0.213	-23.59	-0.116
skaʊk	9.95	0.001	spʌp	11.38	0.182	-1.43	-0.181
sku:k	12.64	0.193	spʌp	11.38	0.182	1.26	0.011
	10.30	0.109		21.26	0.212		

t-test on LND: *t*(14) = 5.26, *p* <.001   *t*-test on CBP: *t*(14) = 5.00, *p* <.001

References

Alderete, John. (1997) Dissimilation as local conjunction. In Kiyomi Kusumo (ed). *NELS 27: Proceedings of the North East Linguistic Society 27* 17--32. Amherst: GLSA.

Baayen, R., Piepenbrock, R. and Gulikers, L. (1995) *The CELEX Lexical Database*. [CD-ROM]. Philadelphia: Linguistic Data Consortium.

Baertsch, Karen and Davis, Stuart. (2003) The split margin approach to syllable structure. In T.A. Hall and S. Hamann (eds.) *ZAS Papers in Linguistics 32* 1—14. Berlin: Zentrum für Allgemeine Sprachwissenschaft, Typologie und Universalienforschung.

Bailey, Todd M. and Hahn, Ulrike. (1998) Determinants of wordlikeness. In Morton Ann Gernsbacher and Sharon J. Derry (eds.) *Proceedings of the 20th Annual Conference of the Cognitive Science Society* 90--5. London: Lawrence Erlbaum.

Bat-El, Outi. (1994) Stem modification and cluster transfer in Modern Hebrew. *Natural Language and Linguistic Theory* 12:571--96.

Berent, Iris, Everett, Daniel L. and Shimron, Joseph. (2001) Do phonological representations specify variables? Evidence from the Obligatory Contour Principle. *Cognitive Psychology* 42:1--60.

Berent, Iris and Shimron, Joseph. (1997) The representation of Hebrew words: evidence from the obligatory contour principle. *Cognition* 64:39--72.

Berent, Iris, Shimron, Joseph and Vaknin, Vered. (2001) Phonological constraints on reading: evidence from the Obligatory Contour Principle. *Journal of Memory and Language* 44:644--65.

Berkley, Deborah Milam. (1994) The OCP and gradient data. *Studies in the Linguistic Sciences* 24:59--72.

Berkley, Deborah Milam. (2000) *Gradient Obligatory Contour Principle Effects*. Ph.D. dissertation, Northwestern University, Chicago.

Boersma, Paul. (1998) *Functional Phonology: Formalizing the Interactions Between Articulatory and Perceptual drives*. Ph.D. dissertation, University of Amsterdam.

Broselow, Ellen. (1991) *The Structure of Fricative-Stop Onsets*. Ms. State University of New York, Stony Brook.

Browne, Wayles. (1981) Slavic -ba and English *slil: two persistent constraints. *Folia Slavica* 4:219--225.

Chomsky, Noam and Halle, Morris. (1965) Some controversial questions in phonological theory. *Journal of Linguistics* 1:97--138.

Clements, George N. and Keyser, Samuel Jay. (1983) *CV Phonology: a Generative Theory of the Syllable*. Cambridge: MIT Press.

Coetzee, Andries W. (2004) *What it Means to be a Loser: Non-Optimal Candidates in Optimality Theory*. Ph.D. Dissertation, University of Massachusetts.

Coetzee, Andries W. (2005) The OCP in the perception of English. In Sonia Frota, Marina Vigario and Maria Joao Freitas (eds.) *Prosodies* 223--45. New York : Mouton de Gruyter.

Coetzee, Andries W. and Pater, Joe. (2006) *Lexically Ranked OCP-Place Constraints in Muna*. Ms. University of Michigan and University of Massachusetts.

Coleman, John and Pierrehumbert, Janet. (1997) Stochastic phonological grammars and acceptability. In *Computational Phonology. Third Meeting of the ACL Special*

33

Interest Group in Computational Phonology 49--56. Somerset: Association for Computational Linguistics.

Davis, Stuart. (1984) Some implications of onset-coda constraints for syllable phonology. *Proceedings of the Chicago Linguistic Society 20* 46--51.Chicago: Chicago Linguistics Society.

Davis, Stuart. (1991) Coronals and the phonotactics of nonadjacent consonants in English. In C. Paradis and J. F. Prunet (eds.) *The Special Status of Coronals: Internal and External Evidence* 49--60. San Diego: Academic Press.

de Lacy, Paul. (2002) *The Formal Expression of Markedness*. Ph.D. dissertation, University of Massachusetts.

Flemming, Edward. (2001) Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18:7--44.

Frisch, Stefan A., Pierrehumbert, Janet B. and Broe, Michael B. (2004) Similarity Avoidance and the OCP. *Natural Language and Linguistic Theory* 22:179--228.

Frisch, Stefan A. and Zawaydeh, Bushra Adnan. (2001) The psychological reality of OCP-Place in Arabic. *Language* 77:91--106.

Fudge, E.C. (1969) Syllables. *Journal of Linguistics* 5:253--86.

Gafos, Adamantios I. (1999) *The Articulatory Basis of Locality in Phonology*. New York: Garland.

Gafos, Adamantios I. (2003) Greenberg's asymmetry in Arabic: a consequence of stems in paradigms. *Language* 79:317--55.

Greenberg, Joseph H. (1950) The patterning of root morphemes in Semitic. *Word* 6:162--81.

Hay, Jennifer, Pierrehumbert, Janet and Beckman, Mary. (2004) Speech perception, well-formedness and the statistics of the lexicon. In J. Local, R. Ogden and R. Temple (eds.) *Phonetic Interpretation: Papers in Laboratory Phonology VI* 58--74. Cambridge: Cambridge University Press.

Hayes, Bruce P. (1985) *A Metrical Theory of Stress Rules*. New York: Garland.

Itô, Junko and Mester, Armin. (2003) *Japanese Morphophonemics: Markedness and Word Structure*. Cambridge: MIT Press.

Itô, Junko, Mester, Armin and Padgett, Jaye. (1995) Licensing and redundancy: underspecification in Optimality Theory. *Linguistic Inquiry* 26:571--614.

Kahn, Daniel. (1980) *Syllable-Based Generalizations in English Phonology*. New York: Garland.

Kawahara, Shigeto, Hajime, Ono and Sudo, Kiyoshi. (2006) Consonant co-occurrence restrictions in Yamato Japanese. *Japanese/Korean Linguistics* 14:27--38.

Kirk, Cecilia. (2001) *Phonological Constraints on the Segmentation of Continuous Speech*. Ph.D. Dissertation, University of Massachusetts, Amherst.

Lamontagne, Greg. (1993) *Syllabification and Consonant Cooccurrence Conditions*. Ph.D. dissertation, University of Massachusetts, Amherst.

Massaro, Dominic W. and Cohen, Michael M. (1983) Phonological context in speech perception. *Perception and Psychophysics* 34:338--48.

McCarthy, John J. (1981) A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12:373--418.

McCarthy, John J. (1986) OCP effects: gemination and antigemination. *Linguistic Inquiry* 17:207--63.

McCarthy, John J. (1994) The phonetics and phonology of Semitic pharyngeals. In Patricia Keating (ed.) *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology 3* 191--233. Cambridge: Cambridge University Press.

McCarthy, John J. and Prince, Alan. (1990) Prosodic morphology and templatic morphology. In Mushira Eid and John J. McCarthy (eds.) *Perspectives on Arabic Linguistics II: Papers From the Second Annual Symposium on Arabic Linguistics* 1--54. Amsterdam: John Benjamins.

McCarthy John J. and Prince, Alan. (1995) Faithfulness and reduplicative identity. In Jill Beckman, Suzanne Urbanczyk and Laura Walsh Dickey (eds.) *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory* 249--384. Amherst: GLSA.

Morelli, Frida. (1999) *The Phonotactics and Phonology of Obstruent Clusters in Optimality Theory*. Ph.D. dissertation, University of Maryland.

Moreton, Elliot. (2002) Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84:55--71.

Newman, Rochelle S., Sawusch, James R. and Luce, Paul A. (1997) Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology* 23:873--89.

Padgett, Jaye. (1995) *Stricture in Feature Geometry*. Stanford: Center for the Study of Language and Information.
35

Pater, Joe. (2005) Learning a stratified grammar. In Alejna Brugos, Manuella R. Clark-Cotton and Seungwan Ha (eds.) *Proceedings of the 29th Boston University Conference on Language Development* 482--92. Somerville, MA: Cascadilla Press.

Pater, Joe and Coetzee, Andries W. (2005) Lexically specific constraints: gradience, learnability, and perception. In *Proceedings of the 3rd Seoul International Conference on Phonology* 85-119. Seoul: The Phonology-Morphology Circle of Korea.

Pierrehumbert, Janet. (1993) Dissimilarity in the Arabic verbal roots. In A. Schafer (ed.) NELS 23: *Proceedings of the North East Linguistic Society 23* 367—81. Amherst: GLSA.

Prince, Alan and Smolensky, Paul. (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*. New Brunswick, NJ: Rutgers University Center for Cognitive Science.

Selkirk, Elisabeth O. (1982) The syllable. In Harry van der Hulst and Norval Smith (eds.) *The Structure of Phonological Representations (Part II)* 337--83. Dordrecht: Foris.

Smolensky, Paul. (1995) *On the Internal Structure of the Constraint Component Con of UG*. Talk presented at UCLA, April 7, 1995.

Sorace, Antonella and Keller, Frank. (2005) Gradience in linguistic data. *Lingua* 115:1497--524.

Suomi, K., McQueen, J. M. and Cutler, A. (1997) Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language* 36:422--44.

Tesar, Bruce and Smolensky, Paul. (1998) Learnability in Optimality Theory. *Linguistic Inquiry* 29:229--68

Tessier, Anne-Michelle. (2004) *Root-Restricted Markedness and Morpho-Phonological Domains*. Paper presented at Montreal-Ottawa-Toronto Phonology Workshop, February 2004, University of Ottawa.

Ussishkin, Adam. (1999) The inadequacy of the consonantal root: Modern Hebrew denominal verbs and output-output correspondence. *Phonology* 16:401--42.

Vitevitch, Michael S. and Luce, Paul A. (1998) When words compete: levels of processing in spoken word recognition. *Psychological Science* 9:325--9.

Vitevitch, Michael S. and Luce, Paul A. (1999) Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40:374--408.

36