International Journal
of
English Studies

IJES

UNIVERSITY OF MURCIA

www.um.es/engphil/ijes

Gradience, Phonotactics, and the Lexicon in English Phonology

MICHAEL HAMMOND*
University of Arizona

ABSTRACT

Experimental work has established that when subjects judge the phonological wellformedness of nonsense forms, they are strongly affected by the frequency of the phonological elements of the form and by the number of actual words that such a form is similar to. These results challenge phonological theory by suggesting a central role for frequency and the lexicon. In this paper, I review these results and show how they can be easily modelled with Probabilistic Optimality Theory. The payoff is that from very few phonological assumptions we can derive virtually the whole panoply of experimental effects. We can also derive various Local Conjunction effects as well.

KEYWORDS: English. phonotactics, Optimality Theory, Gradience, Psycholinguistics, Local Conjunction.

* *Addressforcorrespondence:* **Michael Haniniond, Linguistics Department, University of Arizona, Douglass Building, Tucson, AZ 85721, USA.** E-mail: hammond@U.Arizona.EDU

1. INTRODUCTION

Traditional generative phonology maintains that phonological wellformedness is encoded as a set of rules or constraints. A word is judged illformed if it violates one or more of these rules or constraints. For example, a nonsense form like [bnɪk] is judged illformed because it violates a constraint on what constitutes a wellformed syllable onset of English: stop-nasal sequences are ruled out (Hammond 1999b).

This theory makes two strong and compelling predictions. First, judgments of wellformedness are categorical; nonsense forms are either wellformed or illformed. Second, if a word exists, it is wellformed: the existence of a pattern in the extant lexicon of a language is a sufficient demonstration that that pattern is not ruled out in the language.

Unfortunately, both of these predictions are false. Relevant experimental work shows that in nonsense word tasks, subjects exhibit gradient judgments. For example, a form like [blɪk] is judged as "better" than a form like [sfɪk]. In addition, patterns that should be wellformed by virtue of their existence in the lexicon contribute to the illformedness of nonce forms. For exarnple. though [sf] exists in words like ***sphinx*** or *svelte*, it clearly contributes to the relative illformedness of nonce forms like [sfɪk].

Even more disturbing is that thcse experimental judgments are influenced by occurring similar forms. Thus, a nonce form is judged more wellformed by virtue ofthe number of existing words it is similar to.

In this paper, I first review the basic architecture of generative phonology and orthodox Optimality Theory. I then review the relevant experimental literature, including experiments performed in my own lab. I show how the experimental results present a prima facie challenge, and then 1 show how the results can be handled with a straightfonvard extension of Stochastic Optimality Theory. I show how this extension derives the relevant psycholinguistic properties. I go on to show how various Local Conjunction (Smolensky 1993) effects can also be derived.

II. BACKGROUND

Phonology generally excludes gradience.' Phonological generalizations are categorical and presumably govern a sharp contrast between grammatical and ungrammatical forms. Rule-based phonology maintains that phonological generalizations are described by phonological rules. For example, to capture the fact that voiceless obstruents are aspirated syllable-initially in English, we posit a rule assigning that feature in that environment.

(1) [–vcd] ‣ [+asp] /$_{\sigma}$ [__

To ascertain that a form is consistent with this analysis, we show how by assuming some particular input representation, the analysis produces the required output. ln the case at hand, the only required output is that if a form contains a syllable-initial voiceless obstruent, then that

obstruent rnust be aspirated. For example, if the rule in (1) constitutes the entire rule set, then we can account for aspiration in a forrn likc *take* [t^hek] by showing how we get the right results when we assurne an input likc /tek/. This is shown in the derivation in (2).

(2)

Input:	/tek/
Aspiration Rule:	t^hek
Output:	[t^hek]

To show that a form is *not* consistent with some particular analysis, we show how ***no*** possible input will result in the correct output. This is rather simple in the case at hand, as the analysis is composed of only a single rule. Thus a forrn like [tek] would ***not*** be consistent with the analysis since there is no way such a forrn could ernerge frorn the analysis with an initial unaspirated obstruent.

(3)

Input:	/tek/
Aspiration Rule:	t^hek
Output:	*[tek]

Notice that the analysis as given rnakes no predictions about the wellformedness of forrns like *steak* [stek] or presurnably illforrned *[st^{h}ek]. The rule-based analysis requires aspiration syllable-initially, but does not rule out aspiration in other contexts. To accornplish this, we rnust either add a second rule rernoving aspiration in syllable-medial position or restrict the segrnent inventory. The following rule in (4) irnplernents the forrner idea.'

(4) [–vcd] → [–asp] / $_{\sigma}$ [[+seg]$_1$

The latter alternative is implemented by positing a constraint on input representations. For example:

(5) Input segments are unaspirated.

Either of these analyses out a form like [st^{h}ek]. Under the analysis with (1) and (2), a form like [st^{h}ek] could not escape rule (4). Under the analysis using (1) and (5), a forrn like [st^{h}ek] would violate the constraint on input representations (5).

In either case, notice that there is no gradience: a form either violates the rules and constraints posited or not. A similar situation obtains in orthodox Optirnality Theory (OT).

In OT, the wellformedness of forms is governed by ranked universal constraints. Thus, the requirement that syllable-initial obstruents are aspirated rnight be enforced by a constraint

like the following:

(6) ASPIRATION
Syllable-initial voiceless obstruents rnust be aspirated.

Constraints like *(6)* rnilitate for universal rnarkedness patterns and are in conflict with constraints that enforce a faithful rnapping of input to output forrns.

(7) FAITH
Inputs are identical to outputs.

To capture the fact that aspiration is enforced in a language like English, ASPIRATION rnust outrank FAITH. This is exernplified in (8).

/tek/	ASP	FAITH
☞ a. [t^hek]		*
b. [tek]	*!	

To dernonstrate that a form like [tek], without aspiration, is illformed in English, we rnust show how it is never selected as the optirnal candidate, regardless of what the input form is. In the systern at hand, where there are only the two constraints, ASPIRATION and FAITH, this is fairly simple. For exarnple, if we instead assurne an input form with aspiration, we get exactly the same results.

(9)

/t^hek/	ASP	FAITH
☞ a. [t^hek]		
b. [tek]	*!	

The only difference is that there is now no violation of FAITH for the winning candidate.

The account generalizes to deal with the absence of forms like [st^hek] in a similar way, except that there is only a constraint-based solution. One irnplernentation of this would be to posit a constraint excluding aspiration in all contexts.

(10) NOASPIRATION
Nothing is aspirated

This constraint would be ranked below ASPIRATION, allowing syllable-initial aspiration, but above FAITH, precluding aspiration in other contexts. The following two tableaux show how this works for inputs with and without aspiration respectively.

(11)

/st^{h}ek/	ASP	NOASP	FAITH
☞ a. [stek]			*
b. [st^{h}ek]		*!	

(12)

/stek/	ASP	NOASP	FAITH
☞ a. [stek]			
b. [st^{h}ek]		*!	*

The OT analysis has the same general properties as the rule-based analysis with respect to gradience. Forms are either wellformed or illformed.[3]

III. THE PROBLEM

The problem is that wellformedness judgments are gradient: when presented with nonce forms, subjects can give consistent judgments regarding degree of wellformedness. Moreover, these judgments correlate with two factors: *i)* the frequency with which the pieces of the nonce form occur in real words, and *ii)* the degree of similarity of the nonce form to real words.[4] In this section, I review some of this literature.

For example, Coleman & Pierrehumbert (1997) show that when subjects are asked to rank nonsense forms on a scale of wellformedness from 1 to 7, that their judgments correlate with the frequency of the phonological constituents that make up those forms and with the neighborhood density of those forms.

Frequency of phonological constituents is assessed by breaking forms up into traditional prosodic units. syllables, onsets, rhymes, and then calculating the frequency of those units over a reasonable corpus. The overall frequency score of a nonce form is calculated by multiplying together the frequencies of its sequential parts. For example, the frequency score of a nonsense form like [blɪk] is calculated by determining the frequency of its onset and the frequency of its rhyme and multiplying them together:

(13) $P(blɪk) = P(bl) \times P(ɪk)$

These are multiplied together because of the assumption that they are chosen independently.

Researchers like Greenberg & Jenkins (1964) and Ohala & Ohala (1986) show that neighborhood density also plays a role. Neighborhood density is a psycholinguistic notion that refers to how rnany words a form is "similar" to. The simplest way to do this is to use minimuni edit distance: go through the form frorn left to right, adding, subtracting, or changing the segrnents one by one and add up the nuniber of real words tliat result. This number is the neighborhood density of the form. For exarnple, the neighborhood density of [blɪk] is 13.

(14)

flick	[flɪk]	slick	[slɪk]	click	[k^hlɪk]
brick	[brɪk]	black	[blæk]	bleak	[blik]
bloke	[blok]	Blake	[blek]	blink	[blɪŋk]
blip	[blɪp]	lick	[lɪk]	block	[blak]
bliss	[blrs]				

These researchers show that the greater the neigliborhood density of a nonce form, the greater its wellforrnedness.

Neighborhood density would seern to be sirnply an extreme version of phonotactic probability. That is, if a nonce form shares a phonological constituent with a lot of real forms, its wellformedness goes up as a consequence. On the other hand. if a forrn shares even more material with real forrns, material that may not coniprisc a phonological constituent, then its wellforrnedness also increases, but as a function of increased neighborhood density, rather than phonotactic similarity.

It is possible to disassociate these effects, however. Bailey &Hahn (2001) constructed an experiment where subjects were presented a series of monosyllabic nonsense words that independently varied neighborhood density and phonotactic probability.[5] Items were presented either auditorily or visually and subjects had to rank them on a scale from 1 to 7 for wellformedness.

Bailey & Hahn show that *both* factors play an independent role. That is, we can manipulate neighborhood density and phonotactic probability independently and both factors are significant.

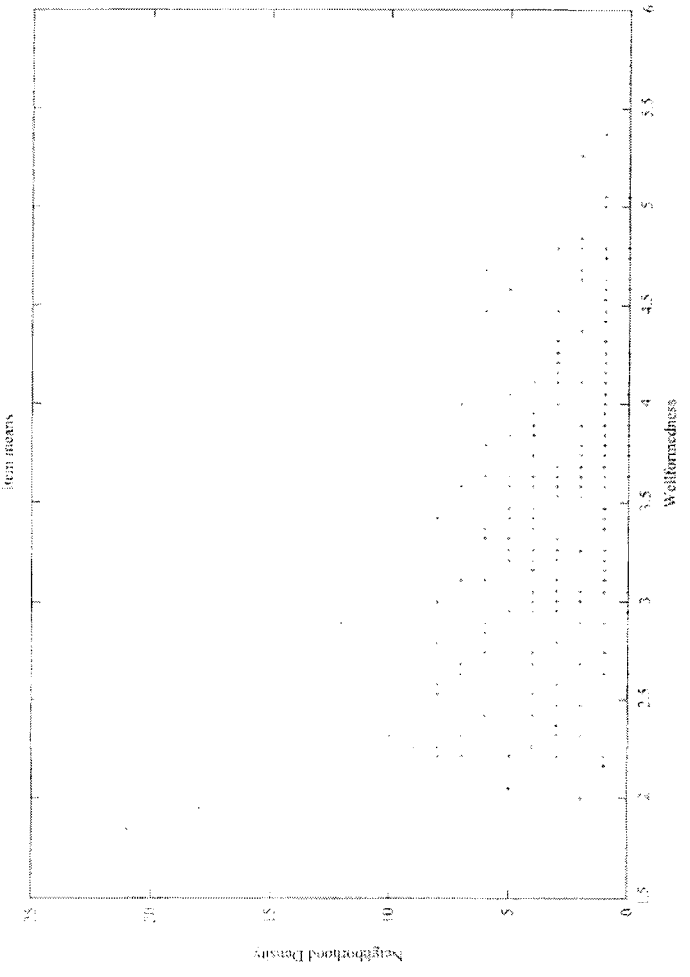
111.1. Replication

This is a powerful and important result and so we undertook a replication of this experiment.[6] Because Bailey & Hahn were able to get the same effects both auditorily and visually, we chose to replicate the visual presentation experiment.' Each subject saw all items and the order of presentation was randoniized for each subject. ln addition, for exploratory purposes, the experiment was run over the web.

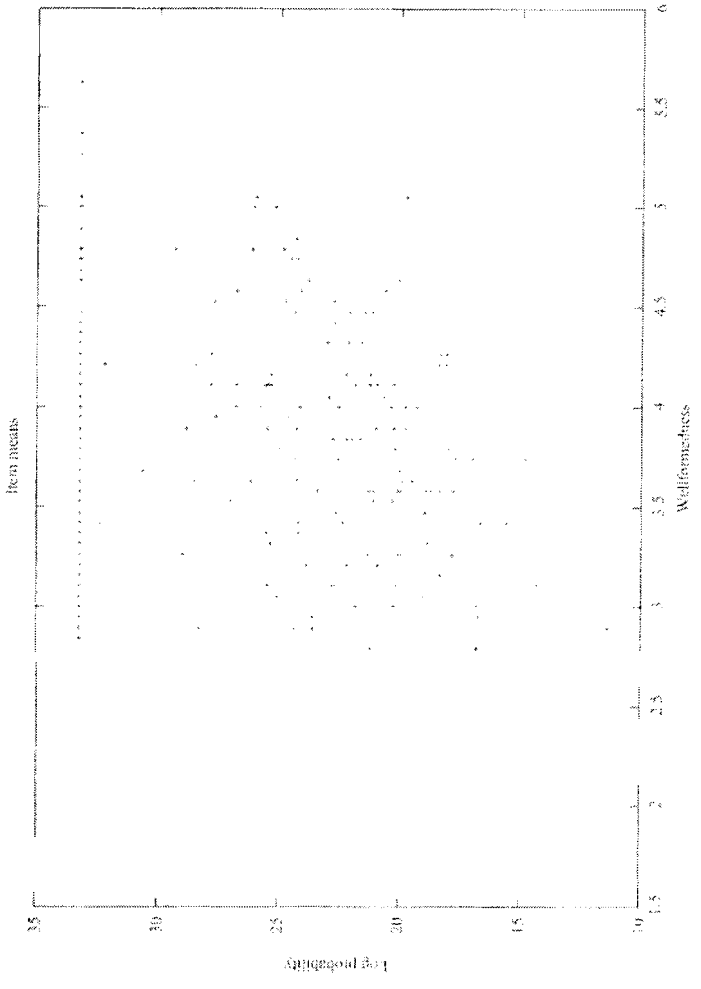
Neighborhood density was calculated as above. Phonotactic probability was calculated by first cornputing the frequencies of onsets and rhyrnes, rnultiplying thern together, and then cornputing the (negative) log probability. This is a standard part of calculating phonotactic probability (Colernan & Pierrehumbert 1997; Frisch *et al.* 2000).

Using a regression analysis, both neighborhood density ($p < .0001$) and phonotactic probability ($p < .05$) had a significant and independent effect on the ratings provided by subjects. As with Bailey & Hahn's results, the effect of neighborhood density was greater than that of phonotactic probability.

The following chart plots itern rneans across subjects against neighborhood density.



The next chart plots item means across subjects against log probabilities.



The upshot is that we confirmed that wellformedness judgments are a function of the frequency with which the components of a nonce form occur in real words and the overall similarity of nonce forms to real words.

III.2. Phonological import

These results pose a challenge to orthodox phonological theory. First, judgments are gradient. We have already seen that traditional rule-based phonology and more recent constraint-based

phonology are based on the assumption that phonological generalizations are categorical.

Second, we have seen that the judgments are a function of phonotactic frequency. Nonce forms composed of more frequent bits are judged more wellformed. Orthodox phonological theory would maintain that if some pattern occurs, then it is wellformed. The frequency with which some pattern occurs should not bear on how wellformed it might be.

Finally, we have seen that there is an independent effect of neighborhood density. The more similar a nonce form is to existing words of English, the more wellformed it is judged. This last effect is an especially troubling result from the perspective of orthodox generative phonology. It is not clear how phonological theory can accommodate overall similarity to existing words.

It might be countered that these effects, though interesting, are irrelevant to phonological theory. Phonology is "about" wellformedness. Wellformedness is assessed through linguistic intuitions and those intuitions are revealed through grammaticality judgments, which are definitionally categorical.

There are two problems with this response. The first is that it simply *defines* judgments as categorical. without empirical or theoretical basis.

The second problem is that we observe gradient effects even when the task is categorical. For example, Frisch *et al.* (2000) replicate the results of Coleman & Pierrehumbert (1997) using both the 1-to-7 rating task. but also using a yes-no task where subjects were simply asked whether a nonce form is wellformed. They got the same effects of phonotactic wellformedness and neighborhood density regardless.

Another possible response is that phonology is not about intuitions per se, but is about something else.[8] It is certainly the case that, in practice, explicit intuitions of wellformedness are not cited as often in phonological work as in other areas of linguistics, e.g. syntax. There are several problems with rejecting intuitions as the empirical base of phonology. however. First, explicit intuitions are cited in phonological research, e.g. Hayes (1984); McCarthy (1982), etc. This sort of work would have to be excluded if the domain of phonology did not include intuitions.

Second. ifphonology is not about intuitions, then what is it about? One possibility might be to claim that phonology is about what we find in language descriptions. The problem with this though is that those descriptions are typically based on the author's intuitions.

Anotherpossibility might be to base phonology on observations in the field. The problem with this, however, is that the set of utterances that occurs naturally is not necessarily an interesting subset —or even a subset!— of the set of utterances that are possible. as Chomsky has argued for decades.

Therefore phonology is about intuitions and needs to accommodate the effects discussed above.

IV. PROBABILISTIC OT

In the following section, I provide an account of these gradient effects in OT. This account builds on a version of Stochastic Optimality Theory and wc therefore review that theory first.

Orthodox OT is built on the assumption that the constraint set must produce a single winner. This assumption can be challenged in several ways and the proposal to be made in thc following section builds on one of these.

As a purely formal matter, the structure of OT does not ***require*** a single output. Specifically, nothing about the theory of constraint interaction necessitates that only a single candidate must win; nothing prevents a tie from resulting. This is shown schematically in (15).

(15)

/abc/	A	...	Z
☞ a. [abc]	*	...	
☞ b. [defl	*	...	
c. [ghi]	**!	...	

This can be established on formal grounds (Hammond 2000a) or on empirical grounds (Hammond 1994).

However, this proposal requires either i) that all languages exhibit multiple outputs in the same contexts, or ii) that not all constraints appear in all languages.

The first scenario is trivial. We must simply allow for some phonological variable that distinguishes candidates that —at least in some context— are not distinguished by any constraints. Overall amplitude might be one such variable. There do not appear to be any constraints that refer to it, but we might choose to encode it in our candidate set. Another more complex possibility might be nasality for glottal segments. While there are a number of constraints that refer to nasality in various contexts, presumably there are none that refer to velum lowering for segments with glottal closure.

The second scenario is a little more complex. Imagine we have two candidate output pronunciations [abc] and [defl for some input /abc/. Imagine further that [abc] and [defl differ only in that some segment in [abc] is specified [+F] and the same segment is specified [-F] in [defl. Then for these to tie in some language, thc constraint enforcing faithfulness to the input with respect to [F] must be absent and any markedness constraints on [F] that might distinguish the candidates must be absent.

An alternative approach is to incorporatc gradiencc into OT by allowing for variable constraint rankings (Anttila 1995). Under this approach, multiple outputs are allowed by leaving some rankings indeterminate. This is represented schematically in (16) where constraints A and B can be ranked n either order (as indicated with thc dottcd linc).

(16)

/abc/	A	B
☞ a. [abc]	*	
b. [def]		*
c. [ghi]	*	*

If A is rankcd above B, [def] is the winning candidate; if B is ranked above A, [abd] is the winning candidate. Noticc that [ghi] loses under cither ranking.

This approacli also allows for niultiple outputs, but does so without giving up on the idea that all languages use the same constraint set.[9]

Anttila proposes tliat when niultiple outputs are possible, their frequency of occurrence corresponds to the nuniber of rankings that produce them. Thus. in the example above. we would expect each output to occur 50% of thetime. Anttila cites a number of more complex cases where more constraints are at play and not all rankings arc distinct. Consider the hypothetical example where tliree constraints are freely ranked, as in (17).

(17)

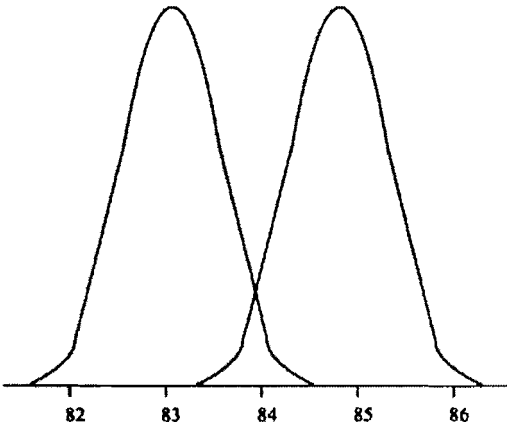
/abc/	A	B	C
☞ a. [abc]	*		**
☞ b, [def]		*	*
c. [ghi]	*	*	**

Here six rankings are possible In four of them —where constraint A or constraint C is top-ranked [def] wins. In the othcr two cases, [abc] wins. Since [def] wins in 4 out of 6 rankings, Anttila's theory would entail tliat it has a relative frequency of .66, while [abc] has a relative frcquency of .33.

A more recent proposal by Boersnia (1997) and Boersnia (1998) proposes that all constraint rankings exhibit a probability value. This is formalized in tlie following way. All constraints are ranked in a Gxed numcrical spacc. For exaniple, in a constraint hierarchy with only three constraints, constraiiit A might exliibit ranking 25, constraint B niight exhibit ranking 51, and constraint C might exhibit ranking 99. The *actual* ranking of a constraint in any particular instance can vary from its numerical position and the set of possible actual rankings for any constraint is describcd by a normal distribution (a bell-shaped curve). These curves overlap and thc likelihood of a ranking inversion is given by the overlap between the probability distributions of the two constraints.

The following picture shows what overlapping normal curves centered at 83 and 85 might look like.

(18)



Hammond (2003b) provides a different formalization of stochastic ranking and we review this here. A set of ranked constraints, A » B » C, can be viewed as a set ofranking relations between each pair of constraints.

(19) Constraints: A, B, C

Relations: A » B
A » C
B » C

Some of these relations can be predicted from others by familiar principles. For example, A » C follows by transitivity from A » B and from B » C.

To accommodate stochastic ranking, each ranking relationship is associated with a probability value (O ≤ n ≤ 1). For example, if there is a one in three chance of A outranking B, we would say P(A » B) = .33. It then follows that the cliancc of the reverse ranking is .66, e.g. P(B » A) = 1 - .33 = .66.

Rankings can be combined on this view in the obvious way. For example, if the relative rankings of B and C are both equally likely, P(B » C) = P(C » B) = .5, then, given that the two pairs of rankings are independent, we can use the multiplication rule to compute their combined probability; the ranking A » B » C has a probability of .167, e.g. P(A » B » C) = P(A » B) × P(B » C) = .33 × .5 = .167.[10]

Notice that this is different from the way the Boersma model works. On that model. ranking relations are not independent as above; the ranking possibilities for some pair of constraints A and B affect the ranking possibilities of B with any other constraint. Consider a situation where A » B » C. Given a normal curve to describe the actual ranking of a constraint, it will be possible for B to outrank A on occasion. That may occur because the actual ranking of A occurs far to the right under its curve and/or because the actual ranking of B occurs far to the left under its curve. (It may be helpful to look back at the picture above to make sense of this.) Notice that in the latter case, C is far less likely to outrank B. Thus rankings are not independent on the Boersma model." We will make use of the pairwise formalization in the remainder of the paper, designating it *Probabilistic OT* to distinguish it from the Boersma model *(Stochastic OT*).

V. A PROPOSAL

Probabilistic OT, as it stands, describes the frequency of distribution in some corpus of multiple outputs of the same input. I propose to extend this to accommodate the experimental gradient grammaticality effects described in section III above.

Specifically, I propose that gradient grammaticality results when some markedness constraint is ranked gradiently with respect to the faithfulness constraints of the language. Let us see how this might work. Imagine we have gradient wellformedness with respect to some markedness constraint, e.g. ONSET. What this means empirically is that subjects would find forms with initial vowels marginally grammatical. For example. they might find a nonsense form like [ɪp] less wellformed than a nonsense form like [bɪp].

Consider now how ONSET might be ranked with respect to the relevant faithfulness constraints. For our purposes, let us take those to be MAX-IO and DEP-IO.[12] There are six possible rankings.

(20)

a. ONSET » MAX-IO » DEP-IO:	[tapa]
b. ONSET » DEP-IO » MAX-IO:	[pa]
C DEP-10 » ONSET » MAX-IO:	[pa]
d. DEP-IO » MAX-IO » ONSET:	[apa]
e. MAX-IO » ONSET » DEP-IO:	[tapa]
f. MAX-lO » DEP-IO » ONSET:	[apa]

Given an input /apa/ and candidates [apa], [tapa], and [pa], each ranking selects the candidate given to the right. Basically, if ONSET is ranked above either of the faithfulness constraints, then the violation is repaired by violating that faithfulness constraint. If it is above both faithfulness constraints, then the lower-ranked of the two is the one violated.

We are attempting to model gradient wellformedness in a context wliere the repair is unknown. That is, for some nonce form, all we know is that a subject finds it grammatical or ungramniatical. If the subject finds it ungrammatical, then we do not know how the subject would *prefer* to pronounce it, since the subject is not asked for this latter information.

In that context, we are dealing with a ranking conîiguration where the markedness constraint could either be ranked below *all* the faithfulness constraints, or below all but one. We can reduce the ranking possibilities to these two because we do not know how the subject would prefer to pronounce the form (if it is illformed). That is, we can view gradient wellformedness as probabilistic ranking of the relevant markedness constraint with respect to the lowest-ranked faithfulness constraint. Considera ranking: FAITH-1 »FAITH-2» ONSET. If the ONSET constraint is ranked below FAITH-2, then onsets are fully optional and a form like [apa] would be judged as wellformed. If the ONSET constraint is ranked *above* FAITH-2, e.g. FAITH-I » ONSET » FAITH-2, then a form like [apa] is judged as illformed. And, of course. ONSET could be ranked above FAITH-I as well, but the experimental task provides no information on this.

If violations of ONSET are assessed gradiently, then the ranking of ONSET with respect to FAITH-2 is probabilistic.

We have shown how probabilistic ranking can be used to formalize gradient wellformedness. However, we have done so schematically with the constraint ONSET. In the experiments performed to date, the relevant markedness constraints are constraints on wellformed sequences of segments or constraints on wellformed onsets and rhymes. Recall the equation in (13) repeated below.

(21) $P(bl\imath k) = P(bl) \times P(\imath k)$

To complete the story, we must assume that there are markedness constraints that correspond to the set of possible onsets and rhymes and that these constraints are probabilistically ranked with respect to the lowest-ranked faithfulness constraint in the language. For example, we might have constraints like *ONSET/[bl] or *RHYME/[ɪk]. Probabilistic ranking of these with respect to faithfulness produces the gradient judgments observcd.

There are several arguments in favor of this proposal.

First, usingprobabilistic ranking to encode gradience *automatically* captures the fact that this gradience correlates with experience. This follows because we can make use of a version oftl-ie *Gradual Learning Algorithm* (Boersma & Hayes 2001). Thc basic idea behind that theory is that constraints are reranked in a gradual way as a function of cxpcrience. Thus, if a constraint against [bl] as an onset is initially ranked above the relevant faithfulness constraint, it will be probabilistically demoted as a function of exposurc to words that contain [bl] as an onset: thus ranking reflects experience and lexical and phonotactic frcqueney.

A second argument for using probabilistic ranking to encode gradience is that it *automatically* captures the multiplicative effect of separate markedness constraints. Recall that

the experimental literature shows that *all* the relevant markedness constraints contribute to the illformedness of any particular nonce form (13 and 21). This follows automatically from the laws of probability theory. The basic idea is that each constraint against some span of the nonsense form is probabilistically ranked with respect to the lowest faithfulness constraint. A form is judged illformed if either or both of those constraints is ranked above the faithfulness constraint. The mathematics behind probability theory tells us that the chance of one or the other (or both) of the constraints outranking the faithfulness constraint is the product of their independent chances of outranking the faithfulness constraint.[13]

A real-world example may help. Imagine we are concerned with who might walk into the room next. We are interestcd in the likelihood that they would not be wearing red and the likelihood that they would not be malc. In other words, how likely is it that the next person who walks in the room will not be wearing red and/or not be male? One way to figure this out is to work out the likelihood that they would *both* be wearing red and be male. Then subtract that from 1. Thus. if the likelihood that somebody would wear red is .3 and the likelihood that they are male is .5, then the likelihood that both are true is .3 × .5 = .15 (assuming these are independent). Therefore the likelihood that at least one of those is *not* true is 1 – .15 = .85.

Let us look now at a linguistic example. Consider constraints against [bl] as an onset and against [ik] as a rhyme: *ONSET/[bl] and *RHYME/[ɪk]. Both ofthese are ranked probabilistically with respect to the lowest faithfulness conslraint. call it FAITH. If the chance of *ONSET/[bl] being outranked by FAITH is *n* and the chance of *RHYME/[ɪk] being outranked by FAITH is *m*. then the chance of both happening is $n \times m$.

(22) If: P(FAITH » *ONSET/[bl]) = *n*, and
P(FAITH » *RHYME/[ɪk]) = m
then: P(FAITH » *ONSET/[bl]) *and*
FAITH » *RHYME/(ɪk) = $ni \times n$

The chance of either one or both of the constraints outranking FAITH is 1 – (m × *n*). It then follows automatically from a probabilistic interpretation of ranking and the assumption that gradient wellformedness is formalized in those terms that wellformedness correlates with the product of the independent ranking probabilities of relevant markedness constraints.

VI. LEXICAL EFFECTS

The account just proposed accounts for gradient wellformedness intuitions that have been reported in the experimental literature and does so in terms of a fairly orthodox interpretation of probabilistic ranking. Wc have not, however, provided an account of the *lexical* effects in these tasks. Recall that Bailey & Hahn find *independent* effects of phonotactic probability and neighborhood density.

One way to account for the lexical effects might be to manipulate faithfulness constraints in much the same way as markedness constraints. That is, we can have specific markedness constraints that are demoted in response to phonotactic experience and specitic faithfulness constraints that are *promoted* in response to lexical experience. Thus, hearing a word like *hrick* [brɪk] demotes markedness constraints prohibiting [br] as an onset and [ɪk] as a rhyme. Likewise, we might assume some sort of faithfulness constraint that has the effect of requiring /brɪk/ be pronounced as [brɪk] and this constraint is (probabilistically) pronioted.

The key to making this achieve the effect we are interested in is that these faithfulness constraints be sufficiently general so that whatever faithfulness constraint is promoted in response to *hrick* would also apply to all its lexical neighbors. Assuming this to be the case, let us see how this obtains the desired result. For an item like *hrick,* we start with relevant markedness constraints ranked high and the relevant faithfulness constraints ranked low.

(23) $\left\{ \begin{matrix} *\text{ONSET}/[\text{br}] \\ *\text{RHYME}/[\text{ɪk}] \end{matrix} \right\}$ ... » FAITH-[brɪk]

Every time the subject is exposed to a word like *hrick,* the relevant markedness constraints are demoted and the relevant faithfulness constraint is promoted.

The wellformedness of a form is a function of how likely it is that the relevant markedness constraints will be outranked by the relevant faithfulness constraints. A niarkedness constraint like *ONSET/[br] is demoted anytime a word with that onset occurs. A faithfulness constraint that has the effect of FAITH-[brɪk] is promoted every time that word (or one of its lexical neighbors) is encountered. Therefore these two constraints will demote and promote at different rates. Hence, the two effects will be distinct.

Support for this approach to lexical neighborhood effects comes from previous work on lexical effects in phonology. First, there is the very earliest work showing tliat phonological constraints can be lexically restricted. For example, Prince & Smolensky (1993) analyze the position of the *–um–* in Tagalog making use of an alignment constraint restricted to that affix.[14]

Second, there is work like Pater (2000) showing that lexically restricted faithfulness constraints are required for a proper treatment of English cyclic stress.[15]

Finally, there is more recent work showing that lexical frequency effects can be encoded with lexically restricted constraints. Hammond (1999a) shows that thc Rhythm Rule exhibits lexical effects. The Rhythm Rule is the phenomenon whereby stress shifts to the left in a modifier depending on the placement of stress in the head noun (Liberman & Prince 1977; Hayes 1984). For example. in isolation, a word like *thìrtéen* has main stress on the second syllable; in combination. the main stress will shift to the left: *thírtèen mén.*

In an experimental study, Hammond (1999a) shows that the frequency of the modifier affects the likelihood of rhythm. Thus relatively frequent modifiers like *abstract* undergo rhythm more readily than relatively infrequcnt modifiers like *abstruse*. This is modeled there with lexically restricted faithfulness constraints, the ranking of which is governed by frequency.'"

Hammond (2003a) argues that reduction in the second syllable of morphologically complex words like *còndènsátion* is a function of the frequency of the whole word *còndènsátion*, but also the frequency of the base *condénse*. The basic idea is that reduction of the second syllable of the derived form is more likely if the form itself is relatively frequent or if its base is relatively frequent. (These effects are statistically independent). Again, this is modeled with differentially ranked faithfulness constraints. where their ranking is determined by frequency.

Thus modelling the lexical neighborhood effect with lexically specific faithfulness constraints both accounts for how the lexical effects are different from the phonotactic effects, but also fits into a range of studies that support this as a mechanism for handling lexical frequency effects.

What is not clear here is how these lexically restricted faithfulness constraints can encode neighborhood effects. Thcre are two possibilities. One is that while the ranking of a faithfulness constraint is affected only by exposure to the lexical item(s) it is specific to, the constraint is interpreted *generously*, so that it controls the faithfulness of all lexical neighbors.

A second possibility is to interpret ihe constraints strictly, but allow their reranking to be affected more generously. That is, a faithfulness constraint is promoted not just when the relevant word is presented, but when any lexical neighbor is presented.

It is not clear whether there is an empirical difference between these two proposals, but the second would seem to be more appropriate at this stage. We have a clear theory of how constraint violations should be assessed and we should therefore be reluctant to accept any weakening of this theory. On the other hand, our understanding of ranking promotions and demotions is still in its infancy.

VII. LOCAL CONJUNCTION

In this section, I show how the proposal developed above accounts naturally for at least some instances of Local Conjunction (Smolensky 1993). This raises the possibility that Local Conjunction can be done away with given the independent need for probabilistic ranking.

Let us consider an example. Hammond (1999b) argues that aparticular pattern in English stress can only be treated with Local Conj unction. The stress pattern of English monomorphemes disallows a secondary stress that is immediately preceded by an unfooted syllable and immediately followed by another stress, ruling out the following four configurations (foot structure is marked with squarc brackets):

(24) ...v] v [Y] [v́...
...v] v [v̀] [Y...
v [v̀] [v́...
v [v̀] [v̀...

This can arise word-initially or whcn the preceding stress is too far to the leít. Thus we find words like [hàma][mèli][dánthc]mum or [Àpa][làchi][cóla] with multiple secondaries. We find words like he[spèri][nós] or a[pèri][tíf] with an initial unfooted syllablc before a secondary not immediately followed by a stress. We also find words like [bàn][dána] and [ràc][cóon] with secondary stress on a degenerate foot.[17] We do not find words like *[Àpa]la[chì][cóla] or *he[spè][nós] and it is therefore these patterns that must be ruled out. The occurring forms above show that this gap is not i) a restriction against secondary stresses, ii) a restriction against unfooted syllables, or iii) a restriction against secondary stress on a degenerate foot.

Hammond (1999b) argues that this rcstriction can be captured by locally conjoining a number of constraints, e.g. PARSE (to avoid the initial unstresscd syllable), *CLASH (to avoid adjacent stresses), and *SECONDARY (to avoid secondary stresses). Thc basic idea is that. while any one of these constraints is not highly ranked enough to rule such a forrn out, they can be combined into a single high-ranked consiraint that can rule these out.[18]

There is another possibility, however, in terrns of the kind of analysis we have been pursuing here. Let us suppose tliat these three individual constraints are each not ranked high enough to rule out these forrns; they are crucially rankcd below some faithfulness constraint." However, they are probabilistically ranked below that constraint. Wliile the chance of any one of thern outranking the relevant faiihfulness constraint is low enough that violations of each occur. The chance of *at least* one of them outranking the faithfulness constraint is high enough so that the combination does not occur.

Let us make this a little more concrete. Assume that the probability that any one of these might outrank faithfulness is .1. If so, then thc probability that at least one of them will outrank faithfulness is ,271 (1 – (.9 × .9 × .9)).

On this view, Local Conjunction as a formal device is not necessary. The Local Conjunction *effect* arises when some set of constraints is probabilistically ranked in such a way that the probability of at least one of theni playing a role is sufficient to have an effect.

Tliis is a more restrictive position than Local Conjunction. It is more restrictive because the Local Conjunction effect is only possible when the individual constraints are ranked sufficiently high so that their combined probability of outranking the relevant faithfulness constraint is significant. Formal Local Conjunction, on thc other hand, simply requires that the simplex constraints be in the system somewhere (bclow the locally conjoined constraint).

This method of deriving Local Conjunction effects is confirmed by the work of Berkley (1994) and Pierrehumbert (1994). Consider this effect discussed by Berkley.[20] Monosyllabic words in English may contain identical obsiruents as in, e.g. pope [p^{h}op], *tat* [t^{h}æt] and *cake*

[k^hek]. However, if the initial cluster begins with an [s], then only identical coronals are allowed, e.g. *state* [stet], but not *[spep] or *[skek]. Berkley argues that this is a statistical effect. It is a consequence of the relative rarity of [s]-obstruent clusters and the relative rarity of non-coronals in final position.

On the analysis proposed here, this emerges automatically from probabilistic ranking. The latter two effects are mirrored by constraints that are probabilistically ranked with respect to the lowest faithfulness constraint. Each one is ranked low enough so that there are occurring forms that violate it. When we consider them jointly, however, the chance of one or the other or both outranking the relevant faithfulness constraint increases to the point that no violations occur.

In our terms, there would be a constraint against [s]+obstruent clusters and a constraint against word-final non-coronals. Thcir *combined* ranking probability with respect to the lowest-ranked faithfulness constraint accounts for the illformedness of forms like *[spep].

VIII. DISCUSSION

We have reviewed Optimality Theory and some of the psycholinguistic literature on wellformedness. At first blush, this literature would seem to pose a problem for traditional phonological theory. Judgments are gradicnt and depend on the frequency of phonological patterns and on the frequency of similar words. It is hard to suppress the urge to dismiss this literaturc as definitionally irrelevant to the concerns of phonologists.

On closer inspection, however, these data can be accommodated quite easily using probabilistic ranking in OT. If wc make the assumption that gradient wellformedness corresponds to probabilistic ranking of markedness and faithfulness, then a number of effects in the experimental literature follow naturally. For example, the fact that wellformedness corresponds to the product of phonotactic probabilities emerges from the basic math of probability theory, as applied to constraint rankings.'' In addition, the fact that gradient wcllformcdness correlatcs with frequency emerges from the Gradual Learning Algorithm (as applicd to Probabilistic OT). the method by which probabilistic rankings are acquired.

The framework also provides a very natural account of the difference between phonotactic effects and neighborhood effects. The former follow from the demotion ofparticular markedness constraints, while the latter follow from tlie promotion of particular faithfulness constraints.

Finally, the framework can also derive some Local Conjunction effects, which raises the possibility of doing away with formal Local Conjunction, and replacing it with probabilistic ranking.

The account proposed offers a phonological treatment of psycholinguistic facts which should give us encouragement that these effects are at least partially in the purview of phonological theory. On methodological grounds. this is a welcome result as well. By showing

that probabilistic OT can account for the experimental wellformedness results we suggest that the relevant experimental techniques can be profitably employed by linguists and that the constrained formalisms of linguistics, e.g. Optimality Theory, may be profitably used by psycholinguists seeking to account for quantitative experimental data.

Acknowledgernents

Thanks to Jordan Brewer, Lynnika Butler, Diane Ohala, Ben Tucker and two anonymous reviewers for useful discussion. All errors are my own.

NOTES:

1.There is, of course, a huge sociolinguistic literature where this is not the case, but we set this aside

2.The subscript indicates that the segment must be preceded by at least one seginent in the syllable (Chomsky & Halle 1968).

3.It is possible to interpret violations in a gradient fashion. so that the less optimal candidates exhibit limited wellformedness in proportion to the kinds of violations that rule them out (see Goston 1998).

4.We will see that. though these sound similar, they can actually be disentangled

5.The items are all given in orthographic form in the appendix.

6.This work was done with Lynnika Butler, Jordan Brewer and Ben Tucker in the SPAM Lab at the University of Arizona.

7.Visual representation using English orthography raises some interesting questions about how subjects decode orthographic representations and whether orthographic factors might play a role when stimuli are presented visually. However, since Bailey & Hahn got the same results visually and auditorily, and because it is so much easier to run the experiment visually, we went with visual presentation.

8.Thanks to an anonymous reviewer for drawing this possibility to my attention

9.Notice that the two approaches are in principle empirically distinct as removing one or more constraints from the hierarchy is not the same as allowing free ranking among some subset of tlie constraints. We set this issue aside.

10.Notice that not all rankings are independent. Thus, in the example at hand, the relative ranking of A and C isnot independent of the rankings of the other two constraints. In this case, this follows automatically, of course, from transitivity of ranking relationships.

11.Thanks to an anonymous reviewer for useful discussion of this point.

12. We assume Correspondence Theory (McCarthy & Prince 1995) for purposes of this paper, but nothing essential hinges on this assumption. MAX-IO forces input elements to appear in the output and militates against deletion; DEP-IO forces output elements to appear in the input and militates against insertion.

13. Recall that this is not quite true in the Boersma model. Any number of markedness constraints can be probabilistically ranked with respect to the lowest faithfulness constraint, but the likelihood of more than one of them outranking that constraint is not the product of their separate reranking probabilities.

14. An anonyinous reviewer points out that there are no VC prefixes, so the lexical restriction may not be necessary.

15. A more extreme position is taken by Russell (1995) and Hammond (2000b) who argue that all lexical information should be encoded by constraint.

16. See Pater (2000) for a similar approach.

17. Note then that a bisyllabic minimum on secondary stress feet will not work.

18. Other constraints playa role in the system as well. For example, there are constraints that allow the expression of quantity sensitivity and lexical stress that allow for the degenerate feet present in the examples cited. These constraints are, of course. outranked by the constrtaint in question. See Hammond (1999b) for more discussion.

19. Faithfulness must be involved to accommodate lexical secondary stresses and to prevent violations from surfacing simply by specifying stress in the input.

20. This is due originally to Davis (1989).

21. A very interesting unresolved question is whether these judgments would correlate with the different ranking probability values provided by Stochastic OT.

REFERENCES

Anttila, Arto (1995). Deriviiig variatioii froin graminar: A study of Finnish genitives. ROA-63.

Bailey, Todd M. & Hahn, Ulrike (2001). Determiiiaiits of wordlikeness: Plionotactics or lexical neighborhoods? *Journal of Memory and Language*, 44, 568-591.

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

Boersma, Paul (1997). How we learn variatioii, optionality aiid probability. ROA-221

Boersma, Paul (1998). *Functional Phonology*. Unpublished doctoral dissertation, University of Amsterdam.

Boersma, Paul & Hayes, Bruce (2001). Einpirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry*. 32, 45-86.

Cliomsky, Noain & Halle, Morris (1968). *The Sound Pattern of English*. New York: Harper & Row.

Coleman. John & Pierrehumbert, Janet (1997). Stocliastic phonological graininars and acceptability. In *Computational Phonology: Third Meeting of the ACL Special Interest Group in Computational Phonology*. Somerset: Associatioii for Computational Linguistics, 49-56.

Davis, Stuart (1989). Cross–vowel plioiiotactic constraints. *Computational Linguistics*, 15, 109-111

Frisch, S., Large, N.R. & Pisoni, D.B. (2000). Perceptioii of wordlikeness: Effects ofsegiiieiit probability aiid length on the processiiig of nonwords. *Journal of Memory and Language*, 42, 481-496.

Golston, C. (1998). Constraint–based inetrics. *Natural Language and Linguistic Theory*, 16, 719-770.

Greenberg, J.H. & Jenkins, J. J. (1964). Studies in the psycliological correlates of tlie souiid systein of Ainerican English. *Word* 20, 157-177.

Haminoiid, Michael (1994). Aii OT account of variability in Walmatjari stress. ROA-20.

Haininond, Micliael (1999a). Lexical frequeiicy and rhythm. In Micliael Darnell *et al.* (Eds.), *Functionalism and Formalism in Linguistics*. Ainsterdam: John Benjamins, 329-358.

Haminond, Michael (1999b). *The Phonology of English*. Oxford: Oxford University Press.

Haminond, Michael (2000a). Tlie logic of OT. *WEC'OL*. 28, 146-162.

Hammoiid, Michael (2000b). Tliere is no lexicoii! *Coyote Papers*, 10, 55-77.

Hammond, Michael (2003a). *Frequency, cyclicity, and optimality.* Iiivited talk, Second Iiiteriiatioiial Koreaii Phonology Coiifereiice, Seoul Natioiial Uiiiversity.

Haininoiid, Michael (2003b). Phonotactics and probabilistic raiikiiig. In A. Carnie, H. Harley & M. Willie (Eds.), *Formal Approaches to Function in Grammar: In Honor of Eloise Jelinek.* Amsterdain: Joliii Benjamins, 319-332.

Hayes, Bruce (1984). The plioiiology of rhythm in Eiiglisli. *Linguistic Inquiry*, 15, 33-74.

Idsardi, William J. (1992). *The Computation of Prosody.* Unpublished Doctoral Dissertation, MIT.

Liberinaii, M. & Priiice, A. (1977). On stress aiid linguistic rliytliin. *Linguistic Inquiry,* 8, 249-336.

McCarthy, Joliii (1982). Prosodic structure aiid expletive infixation. *Language.* 58, 574-590

McCarthy, Joliii & Priiice, Alaii (1995). Faithfulness aiid reduplicative ideiitity. In J. Beckinaii, L. Dickey & S. Urbaiiczyk (Eds.). *Papers in Optimality Theory. U. Mass. Occasional Papers in Linguistics*, 249-384.

Oliala, Joliii & Oliala, Manjari (1986). Testiiig Iiypotlieses regardiiig tlie psychological inanifestatioii of morpheme structure coiistraiiits. In Joliii Oliala & Jeri Jaeger (Eds.). *Experimental Phonology.* Orlaiido: Acadeinic Press. 239-252.

Pater, J. (2000). Non-uniformity iii Eiiglisli secoiidary stress: tlie role of raiiked aiid lexically specific coiistraiiits. *Phonology,* 17, 237-274.

Pierreliuinbert, Jaiiet (1994). Syllable structure aiid word structure. Iii Patricia Keatiiig (Ed.), *Pupers in Laboratory Phonology III.* Cninbridge: Cainbridge Uiiiversity Press, 168-190.

Priiice, Alaii & Sinoleiisky, Paul (1993). Optiinality Tlieory. Ms, U. Mass aiid U. Of Colorado

Russell, Keviii (1995). Morpheme aiid candidates iii Optiinality Tlieory. ROA-44.

Sinoleiisky. Paul (1993). Harinoiiy, iiiarkediess aiid phonological activity. Paper preseiited at tlie Rutgers Worksliop 1.

APPENDIX

Materials froin Bailey & Hahn used in the replication reported liere.

drolf	greltch	prunth	slirupt	smisp	stolf
threlth	zinth	drusp	gwesht	shandge	slesk
snulp	swesk	tliroiidge	glemp	krenth	shresp
slontch	spulsh	swust	trinth	binth	drilf
glump	inisp	shadge	slintch	spulp	thrilth
bleinp	drisp	greld	nulp	shan	slisk
spulse	thriiidge	blesk	droff	grell	pinth
shendge	slisp	spust	tlirond	breltch	drosp
grelm	plemp	shesp	slon	stelf	throiiii
brelth	druinp	grentch	plunth	shindge	slotch
stilf	thrupt	brenth	drup	gresht	presp
shinth	slulp	stoff	tlirusp	bresp	drupt
gresp	printh	shondge	sluntch	stulf	tolf
brondge	drusk	gretch	prolf	shrapt	slusk
stulp	treltch	bruiitli	druss	griltch	pruiidge
shremp	slust	stust	trenth	cleinp	drust
grinth	prunt	slirep	siniinp	sulp	tresp
clenth	dusp	grolf	pruntch	shrept	smip
sulsh	trilth	clontch	dweslit	groiidge	prupt
shress	smiss	sweck	trin	cren	dwesk
grunth	punth	shrest	smist	swelk	triiidge
crend	finth	grupt	quentli	shript	smust
swesht	trintch	crendge	flemp	grusp	reltch
shrisp	snalp	swess	trith	crent	flesk
qweft	relth	shruct	snisp	swest	trolf
crentch	flontcli	qwelt	renth	shruft	snulf
swisk	trupt	cresp	freltch	qwept	resp
shrunt	snulk	swisp	trusp	creth	freiith
gwesh	rinth	shrup	snull	swist	twinth
crinth	frinth	gwet	rolf	shrust	snult
swontch	wesht	crondge	froiidge	hinth	rondge
shrut	snump	swuft	wesk	cruiitli	frunth
inth	rupt	shuiidge	siiup	swunt	winth
crupt	frupr	jinth	rusp	sinth	snust
swupt	wust	crusp	frusp	kenth	sandge
sisp	solf	swutt	yinth	dinth	geltcli
kinth	scolf	skisp	sontch	threll	zilth
dolf	gemp	kwesk	scontch	sleck	spelsh
threlm	zin	drelf	gesht	lemp	sculp
sleinp	spesk	threlsh	zindge	dreltch	glein
lesk	sculsh	slentch	spuldge	threlt	ziiit
dreiitli	glep	linth	scust	sless	spulk
tlireth	zintch	dresp	glimp	minth	sesk
slest					