ISCA ITRW on
Spoken Word Access Processes (SWAP)
Nijmegen, The Netherlands
May 29-31, 2000

WHY PHONOLOGICAL CONSTRAINTS ARE SO GRANULAR

Janet Pierrehumbert
Northwestern University, Evanston IL

ABSTRACT

Speech perception models with a fast phonological preprocessor (an FPP) require a uniform phonology throughout the speech community, because chunking errors would otherwise impede lexical access. The phonological constraints encapsulated in an FPP arise as regularities over the lexicon. However, different people know different words. A Monte Carlo simulation of vocabulary acquisition using the Celex database reveals that coarse-grained phonological constraints can be acquired reliably from 30% or less of the monomorphemic words (e.g. by early adolescence). An hypothetical fine-grained constraint combining a trisyllabic stress template and a nasal-obstruent restriction cannot be reliably learned without the large vocabulary of an educated adult. Thus, the nexus of speech perception, phonological well-formedness, and individual variation in vocabulary means that phonological constraints must be coarse-grained.

1. INTRODUCTION

Results of experiments on lexical access during speech perception support cognitive models in which the mental lexicon is characterized as a network. Words are incrementally activated as the speech signal is perceived over time. A word is recognized when its overall activation level passes threshhold, with the time point at which this event occurs depending on the baseline activation, the degree of similarity between the speech signal and the lexical representation, and the extent of competition from other similar-sounding words.

Models of this general class are divided with regard to the status of "phonology", or general implicit knowledge of the sound patterns of a language. The pioneering TRACE model [1] locates phonology in situ. The model has no level of representation which encodes phonology as such and which is distinct from the lexicon. The phonological system is treated as epiphenomenal. A second class of model, typified by SHORTLIST [2], MERGE [3], and NAM [4], has a fast phonological preprocessor (hereafter, an FPP) which uses language-specific but still general prosodic and phonotactic patterns to chunk the speech stream on its way up to the lexical network.

The evidence for models with an FPP includes an experiment by Pitt and McQueen [5], showing that phonotactic patterns impact speech perception even after the unique point of a word. A series of experiments by Vitevich, Luce, and colleagues (reviewed in [4]), shows that the interaction of lexical neighborhood effects and effects of phonological pattern frequency is heavily task dependent. Models like [1] have difficulty explaining the decoupling of these effects, because they attribute both to a single level of representation.

Results to date permit some inferences about the relationship of the FPP to the phonological system as conceived by linguistics. The FPP appears to be sensitive to the inventory of phones and/or phonemes of the language. It reflects syllable and foot structure, as well as junctural probabilities. In short, it implicitly encodes constraints on well-formedness. There is no evidence that it encodes patterns of morphophonological alternation. Morpho-phonological alternations emerge in models of this class as generalizations over related lexical items. If they are encoded in any separate level of representation, this level would be distinct from the FPP.

An important question for psycholinguists is what types of phonological patterns are candidates for being encoded in the FPP. For linguists, the comparable question is what types of observable regularities are candidates for inclusion in the grammar. I would like to explore the causes and consequences of a simple answer to these questions, namely that the constraints are coarse-grained. That is, they lack logical complexity and can be represented as relatively small fragments of phonological structure, which are relatively lacking in detail. In particular, familiarity with the linguistics literature suggests a tradeoff between temporal scope and phonemic detail which points to restrictions on overall granularity. Phonological constraints which refer to specific phonemes, such as idiosyncratic constraints on syllable onsets, typically involve only extremely local phonological descriptions. Constraints at a larger temporal scale, such as vowel harmony rules and word-level stress placement, tend to involve broad classes of phonemes if they refer to phonemic properties at all.

It may be difficult to realize that language could in principle be otherwise. However, astonishingly detailed cognitive representations have been demonstrated in other areas, such as memories of photographs or sequences of events. Mental representations of individual words are far more complex than phonological constraints appear to be. Pitt and McQueen's reasoning in [5] in fact depends critically on the implicit assumption that the phonological window supporting implicit knowledge of probabilities is smaller than the whole word. Why are phonological constraints so much less complex than cognitive representations generally?

The proposal which will be advanced here depends on the tight connection between phonology and processing, as posited in models with an FPP. That is, it depends on the assumption that phonology impacts not only well-formedness judgments, but also the exact patterns of allophony in production and the exact strategies of chunking in perception. It follows that the different speakers of a language community need to have identical or highly similar phonologies. If a speaker and a listener had different phonologies, then the allophonic patterns of the speaker could mislead the listener about how to chunk the speech stream. Such a mismatch in structuring of the speech would impede lexical access.

Clearly, people do not all know all the same words. Effective communication can occur between people with different vocabularies because unknown words can be segmented out with reference to known words and their meanings can be established by context or further discussion. The hypothesis explored here is that phonology is more unanimous than vocabulary. I undertake to explain the fact that the phonological constraints are coarse-grained with reference to the claim that the learning of phonology needs to be robust across variations in individual vocabularies. Phonology is shaped by ecological validity and overly detailed phonological grammars are not viable in a diverse speech community. (See [6] for a similar argument about morphology). I take no stand on whether nonviable grammars are eliminated during learning, or are precluded by constraints on cognitive complexity which have arisen through evolution. What concerns me here is establishing the necessary, mathematical, connection between constraint granularity and vocabulary variation.

2. METHODS

The results presented here are calculations over phonological transcriptions and word frequency data from the Celex English database. A list of 11382 monomorphemic words in Celex was compiled by linguistics graduate students working in the Ohio State University phonetics laboratory. This list includes words which are entered in the database as monomorphemic, and words which are entered as having obscure morphology but were judged by all three students to be nondecomposed. It excludes words which appeared to be formed on a semi-productive reduplicative compounding pattern of English (words like "willy-nilly"). Given the existence of a few errors in Celex and the subjectiveness of judgments used to compile the list, the list is perhaps better characterized as words of low decomposibility. This list figures in the study as the community vocabulary of nondecomposible words whose nature determines what should receive a unitary parse by the FPP.

Individual vocabularies of various sizes were constructed from this list by frequency weighted downsampling of the community vocabulary set. Vocabularies of five different sizes were computed: 400, 800, 1600, 3200, and 6400 words, representing individuals at different levels of vocabulary development. Words of count 0 in Celex were assigned a count of 1. For each vocabulary size, many different vocabularies were generated through independent random samples of the full set. The downsampling is frequency-weighted because it is more likely that a child would learn a frequent word than an infrequent one. The result is that a frequent word, such as "child" shows up in many more vocabularies than a rare word such as "hake".

The assumption that the likelihood of acquiring a word is linear with frequency may be an oversimplification; if a word must be encountered several times to be acquired, then the likelihood would correspond to a power of the frequency. The disregard of semantic factors, however, is not particularly damaging to the present enterprise, because phonological constraints are not particularly correlated with semantic structure. The aggregate phonological statistics of the smaller lexica are probably illustrative even if their semantic structure is not.

The following lists show 15 representative selections from vocabularies of each of the five different sizes. The words are shown together with Celex counts for its 17.9 million word database.

400 WORDS afraid (2012) ambition (432) danger (1695) child (19389) sea (3102) magnificent (351) claw (162) demand (1433) voice (4823) almost (8426) stir (735) candle (294) leave (14042) body (6520) idea (7132)

800 WORDS mean (13126) garden (2517) million (3518) barrel (379) farm (1530) explain (3606) weather (1164) spirit (1362) broad (911) process (3194) swallow (594) deed (180) harry (446) sum (864) part (10237)

1600 WORDS honest (644) annoy (315) oocupy (911) anus (38) examine (1387) tube (428) drag (867) maintain (1713) enthral (23) wine (1421) stir (735) crisis (1181) ghost (554) polite (390).

3200 WORDS vaccine (66) capon (17) mine (1426) centenary (11) vinegar (157) purr (59) flick (190) arrest (601) simultaneous (70) tray (459) shiver (310) west (619) barge (85) harp (50) glory (430)

6400 WORDS twine (27) havoc (67) arena (118) capsize (18) fanatic (22) excrete (19) digest (151) refract (11) wreak (24) hake (6) jeopardy (32) truant (35) dike (48) molest (28) hem (56)

A preliminary estimate of the age level corresponding to each word set indicates that early adolescence corresponds to the 1600 or possibly the 3200 word level. (Recall that this does not represent the total vocabulary, only the vocabulary of nondecomposible words.) This is a an important reference point for theories of language acquisition with a critical period, and the acquisition of phonology is believed to be largely complete at this time. A 6400 word list includes many words which would be known only by educated adults. It would be problematic to suppose that learning phonology critically depends on having such an extensive vocabulary.

Three different classes of phonological regularities were evaluated for each vocabulary size. The most simple is an extended form of the basic trochaic pattern of English. As is well known, trochaic disyllables are far more common than iambic disyllables in English. This regularity is ambiguous in its extension to trisyllabic words. A trochee positioned at the left edge of a trisyllable results in a 100 stress pattern. Phonological theory describes such as pattern as involving left-to-right stress assignment and/or final extrametricality. If a trochee is positioned at the right edge of a trisyllable, a 010 pattern results, indicating right-to-left stress assignment and/or initial extrametricality. Since languages differ typologically on exactly this dimension, the relative well-formedness of these patterns must be learned from language exposure.

The second phonological regularity evaluated was a set of junctural constraints on a nasal in coda position followed by an obstruent in onset position, (hereafter, NO clusters). This set was chosen because a set of experiments by Hay, Pierrehumbert, and Beckman [7] showed that the frequencies of such clusters in the lexicon are gradiently reflected both in speech perception and in well-formedness judgments. Of the clusters used in these experiments, five were selected for analysis here. These are: /n.t/, /n.s/, /n.f/, /m.f/ and /n.p/. These five clusters fully span the range of likelihoods, with /n.t/ being the most frequent NO medial cluster and /n.p/ being entirely absent in monomorphemes. /n.s/ is quite frequent and is judged to be highly well-formed, though not quite as highly as /n.t/. /n.f/ represents a cluster which is not common, but is nonetheless still more common within monomorphemes than across a word boundary. As explained in [7], the monomorphemic parse still wins statistically over a parse which imputes a morphological decomposition, and accordingly the FFP should not posit a boundary. /m.f/ occurs in only 12 monomorphemic words in English. A decompositional parse wins statistically over the monomorphemic parse and dictates the well-formedness judgment. Nonetheless, speakers may still have implicit knowledge that this cluster is a possible one, in the sense that a monomorphemic loan word containing this cluster could probably be added to the English lexicon without reanalysis. /n.p/ is completely impossible except across a word boundary.

The data in [7] indicate that learning the constraints on NO clusters means learning the complete ranking of these five types. Some other clusters, such as /n.s/ and /m.p/, are so close in likelihood that speakers may not distinguish them in terms of well-formedness. Note also that the ranking /n.t/> /n.s/ > /n.f/ > /m.f/ > /n.p/ is not the same as the nasal homorganic rule one might find in a linguistics textbook. /n.s/ perceived to be better than /m.f/ even though both are homorganic fricative combinations. /n.f/ is better than /m.f/ even though /n.f/ is nonhomorganic. This is an example of how a detailed experimental study may show that the observations in a linguistics textbook are overly coarse compared to the actual state of affairs.

The third phonological pattern evaluated was a statistical pattern which accurately describes the monomorphemes as a group but which appears excessively detailed as a candidate phonological constraint. This is an hybrid of the first two targets of investigation; a set of NO constraints specific to trisyllabic words with initial stress. The constraint set involves the same clusters as in the second calculation, but these clusters are now confined to the juncture between the first syllable, which is stressed, and the second syllable, which is unstressed. The ranking of the clusters is the same in the total monomorphemes set as the ranking for the five NO clusters without a stress constraint.

The relative learnability of each of the three target patterns was computed as follow: 20 independent random samplings of the community vocabulary were drawn for each of the five vocabulary sizes. For each individual vocabulary, the counts of the target patterns were established. The counts were converted to ranks, with individuals viewed as agreeing in their grammars if they agreed in the ranking. If all twenty individuals have the same ranking, then the pattern in question is obviously learnable with $P < 0.05$ of any error. Any disagreement in ranking means that agreement among 20 individuals on that ranking cannot be assured, given the vocabulary size.

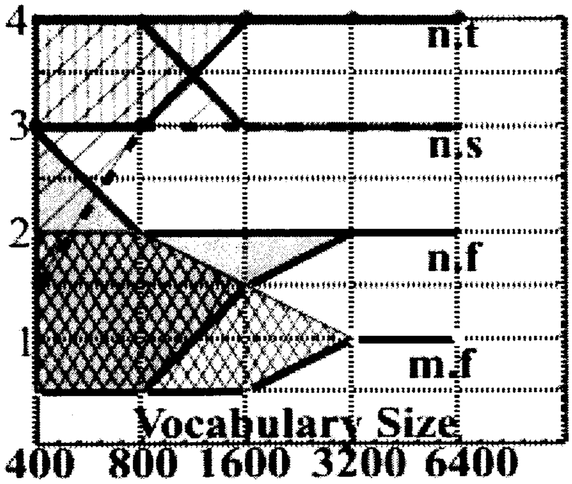
The two measures of failure that will be of interest here are the total range of rankings for any given stress or NO template, and the Spearman rank correlation of the individuals' rankings to the true rankings in the total word list. If the size of the total range of ranks for some particular template is greater than zero, then at least some individuals failed to learn the same ranking as others. For example, in the NO calculations for vocabulary size 800, some individuals have /n.t/ ranked as 4 (the top rank, above /n.s/) and others rank it 3 (below /n.s/). These latter individuals had a different grammar from those who ranked it best. The mean Spearman rank correlation provides an aggregate measure of how well the different individuals learned the phonological patterns. This number is positive if any learning at all occurred, and it tops at 1.0 for unanimously perfect learning.

3. RESULTS

The relative well-formedness of the 100 stress pattern versus the 010 stress pattern is perfectly learned by all individuals for all five vocabulary sizes. Clearly, this is a highly robust bit of phonology; although it involves a trisyllabic template, it has no phonemic specifications.

A detailed plot of the max-min data for the NO data is presented in Figure 1. The horizontal axis on this figure is the vocabulary size. The vertical axis represents observed rank. Maximum and minimum ranks for all five NO clusters are shown superimposed on the graph. Maximum ranks for the same cluster are connected by a line, as are minimum ranks for the same cluster. The improvement in learning for a given cluster as the vocabulary size increases is represented by a convergence between the maximum and minimum lines for that cluster, with the lines converging to the same value if all individuals learn the same ranking. Each such wedge in the figure is hatched or shaded. A visual impression of the extent of learning for all NO clusters in the set can be gained by examining the relation of these wedges to each other. Any contact between two wedges represents a learning error by at least one individual. Actual overlap between wedges -- shown with overlaid shading -- represents a more severe breakdown of grammatical unanimity.

Figure 1: Evolution of NO ranks as vocabulary size increases.



The unanimity for vocabulary size 400 is very poor for this phonological pattern. By vocabulary size 3200, however, all individuals have the same ranking. It is very interesting to note that this unanimity emerges when individuals have learned only 28% of the community vocabulary. This result strongly supports the suggestion that individuals may have learned the same phonological grammar despite having different vocabularies. The finding appears to be consistent with the idea that core phonological constraints would be in place near the beginning of adolescence.

Table 1 presents the mean Spearman rank correlation as an aggregrate measure of learnability. It compares the learnability of the 100/010 stress contrast, the NO pattern, and the stress-plus-NO pattern.

Table 1: Mean Spearman rank correlations

	Vocabulary size				
	400	*800*	*600*	*3200*	*6400*
Stress	1.00	1.00	1.00	1.00	1.00
NO	0.80	0.88	0.98	1.00	1.00
Stress +NO	0.21	0.45	0.61	0.68	0.93

The stress contrast is completely learnable for all vocabulary sizes. The NO pattern displays a rather good degree of learnability (at 0.8) for even the 400 word vocabulary, and reaches 1.0 at 3200 words, as just discussed. The learnability of the stress-plus-NO pattern is extremely poor for a small vocabulary, and even at 6400 words unanimity has not been reached. With 6400 monomorphemic words representing a high vocabulary level which many adults fail to achieve, it is clear that this regularity is not viable as a shared constraint in the phonology of the community.

4. DISCUSSION

The calculations just presented bear out the claim that phonological regularities can be uniformly learned by individuals with different vocabularies. A rather complex regularity mentioning specific phonemes -- a five way ranking of NO clusters -- can be uniformly learned from 3200 words or less. A stress constraint involving a large temporal scale but no specific phonemes can be learned even more readily. The calculations also indicate that a hypothetical pattern combining a large temporal scale and specific phonemic configurations requires something approaching a full community vocabulary to be reliably learned. Since many people do not have such a large vocabulary, it is not statistically stable enough to be viable in the grammar. Thus, the relationship of statistical reliability to complexity and vocabulary size appears to be exactly in the right range to explain the granularity effects observed.

These calculations exemplify a specific line of reasoning, and do not provide anything like an exhaustive survey of the relationship between granularity and learnability. The calculations presented in [7] in fact presupposed without argument a somewhat finer granularity than used here. Noise in the experimental data make it difficult to determine whether this detail assisted the fit by modelling knowledge more accurately, or impeded it by introducing statistic instability. More detailed experiments are needed to resolve this issue. The results of these calculations also suggest followup work in which individual variation in well-formedness judgments and speech perception is related to assessment of individual vocabularies.

5. ACKNOWLEDGMENTS

Janet Pierrehumbert, Dept. of Linguistics, Northwestern University, Evanston IL 60208, USA. Email jbp@northwestern.edu

6. REFERENCES

[1] McClelland, J.L. & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.
[2] Norris, D. (1994) Shortlist: a connectionist model of continuous speech recognition. *Cognition*, **52**, 189-234.
[3] Norris, D.G., McQueen, J.M. & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, **23**.
[4] Vitevich, M. & Luce, P. (in press). Probabilistic Phonotactics and Neighborhood Activation in Spoken Word Recognition. *Journal of Memory and Language.*
[5] Pitt, M.A. & McQueen, J.M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, **39**, 21-46.
[6] Johansson, C. (1997). *A View from Language; growth of language in individuals and populations.* Ph.D dissertation, University of Lund.
[7] Hay, J.H., Pierrehumbert, J.B. & Beckman, M.E. (forthcoming). Speech Perception, Well-formedness, and the Statistics of the Lexicon. *Papers in Laboratory Phonology VI*, Cambridge University Press.