LANGUAGE AND COGNITIVE PROCESSES, 2001, *16* (5/6), 691–698

Why phonological constraints are so coarse-grained

Janet Pierrehumbert
Department of Linguistics, Northwestern University, Evanston, USA

Current models of speech perception are divided with regard to the status of ''phonology'', or general implicit knowledge of the sound patterns of a language. In the TRACE model (McClelland & Elman, 1986) the phonotactic and prosodic constraints of phonology are treated as epiphenomenal from regularities in the lexicon. In contrast, Norris (1994), Vitevich and Luce (1998) and Merge (Norris, McQueen, & Cutler, 2000) respond to a growing body of experimental literature indicating that low-level encoding of the speech signal (the level whose result is passed up to the lexicon for potential word matches) is sensitive to phonotactic and prosodic constraints. Here, I will explore the consequences of the assumption that the architecture of the speech perception system includes 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. By integrating such information, the FPP imputes possible word boundaries to particular temporal locations in the speech signal. An important question is what types of phonological patterns are candidates for being encoded in the FPP: Can the processing system exploit any statistical regularities whatsoever in the shape of words? Does absolutely any structural description which is logically possible in phonology provide a usable constraint? I will explore the causes and consequences of one observation about this issue, namely that viable constraints are coarse-grained. Although the logical apparatus of phonological theory would make it possible to state extremely fine-grained constraints (e.g., constraints containing complex and detailed combinations of features and structural positions), the constraints for which we now have linguistic or psycholinguistic evidence are considerably simpler.

Requests for reprints should be addressed to Janet Pierrehumbert, Dept. of Linguistics, 2016 Sheridan Road, Northwestern University, Evanston, IL 60208, USA.
Email: jbp@northwestern.edu

http://www.tandf.co.uk/journals/pp/01690965.html DOI: 10.1080/01690960143000218

The explanation advanced here depends on the tight connection between phonology and processing. That is, phonology impacts not only well-formedness judgements, but also patterns of allophony in production and strategies of chunking in perception. It follows that the different speakers of a language community need to have identical or highly similar phonologies. Otherwise, the allophonic patterns of the speaker would mislead the listener about how to chunk the speech stream, and nonrecoverable errors in lexical access would result. However, people do not all know all the same words. I will argue 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. Phonological constraints must be coarse-grained because complex and detailed phonological descriptors are statistically unstable across differences in vocabulary, and cannot be learned reliably. This paper in no way exhausts the issue of statistical robustness, which has many other repercussions for learning; for example, see Johansson (1997) for repercussions in the area of morphology. Similarly, I take no stand on whether the coarse-grained character of phonology is innate or emergent in the cognitive system.

METHODS

The results presented here are calculations over phonological transcriptions and word frequency data from the CELEX English database (Baayen, Piepenbrock, & Gulikens, 1995). A list of 11,382 monomorphemic words in CELEX was compiled by linguistics graduate students working in the Ohio State University Phonetics Laboratory. It includes words which are coded in CELEX as monomorphemic, and words which are coded as having obscure morphology, but were judged by all three students to be nondecomposed. 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. The monomorphemic word set, as opposed to the full dictionary, is used because CELEX includes a large number of morphologically complex words and compounds (such as *bachelor's degree*) which have acquired specialised meanings. Such forms have internal word boundaries and should be excluded from the training set which defines constraints on simplex forms.

Individual vocabularies of various sizes were constructed from the list of CELEX monomorphemes by downsampling of the community vocabulary set. The downsampling is frequency-weighted because it is more likely that a child would learn a frequent word than an infrequent one. However, idiosyncratic experiences can permit a child to learn a rare word that few other children know. Words of count 0 in CELEX were assigned a count of

1, as otherwise they are unavailable as candidates for selection. Vocabularies of five different sizes were computed: 400, 800, 1600, 3200, and 6400 words, representing individuals at different levels of vocabulary development. For each vocabulary size, 20 different vocabularies were generated through independent random samples of the full set.

The maximal sample size computed was 6400 words, because a preliminary estimate of the age and educational level corresponding to inventory size suggested that only rather well-educated adults would know this many morphologically simplex words. This estimate was made by sending random samples of the word lists to two reading specialists. Very small vocabularies are of interest because even young toddlers have significant success in using string-spotting and language-particular phonological patterns to decompose the speech stream (see review in Jusczyk, 1997). The astonishing extent of children's early success in parsing speech should not, however, obscure the fact that many years are required to reach adult levels of performance. Error rates in segmental perception and production tasks are above adult levels until age 7 to 9 (see review in Barton, 1980). Allophony, stress/accent, and phonetic precision continue to develop from 6 to 12 (Atkinson-King, 1973; Chevrot, Beaud, & Varga, in press; Eguchi & Hirsh, 1969; Kent & Forner, 1980; Lee, Potamianos, & Narayan, 1999; Raimy & Vogel, 2000).

Four different phonological regularities were evaluated for each vocabulary size. (1) An extended form of the basic trochaic pattern of English. Results in Cutler and Butterfield (1992) show that the predominance of words with initial stress in English is exploited in speech perception to hypothesise word boundaries. Phonologically, this means that English has a trochaic foot structure (a strong-weak foot structure), and that a foot is positioned at the left edge of most words. A trochee yields a 100 stress pattern if positioned at the left edge of a trisyllable, and a 010 stress pattern if positioned at the right edge. Since languages can differ typologically on both foot structure and foot alignment (see Hayes 1994), both must be learned to yield the experimental findings. Here, I evaluate the learnability of the foot alignment for trisyllables, e.g., the learnability of the generalisation that 100 stress patterns are better than 010 stress patterns in English.

(2) 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). A set of experiments in Hay, Pierrehumbert, and Beckman (in press) 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 Hay et al. (in press), five were selected for analysis: /n.t/, /n.s/, /n.f/, /m.f/ and /n.p/. These span the range of cluster frequencies. /n.t/ is the most frequent

NO medial cluster. /n.s/ is quite frequent and is judged to be highly well-formed, though not as highly as /n.t/. /n.f/ is not common, but is nonetheless more common within monomorphemes than across a word boundary. Accordingly, the FFP should not posit a word boundary. /m.f/ occurs in only 12 monomorphemic words. A decompositional parse wins statistically over a simplex parse and dictates the well-formedness judgement. Nonetheless, speakers may still have implicit knowledge that this cluster is a possible one. 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.

(3) The third phonological pattern evaluated was a statistical pattern which accurately describes the monomorphemes as a group but which appears excessively detailed as a phonological constraint set. This is a 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 (in actuality) the same in the total monomorphemes set as the ranking for the five NO clusters without a stress constraint. If such a complex constraint were learnable, then it would be possible for a language to have a phonological grammar with co-occurrence restrictions for NO clusters which depended on the stress pattern of the word. Such complex constraints are not reported.

(4) Lastly, I evaluate the learnability of a regularity discussed in Moreton (1997): Word-final stressed /gri/ (as in *agree*) has higher frequency than word-final stressed /kri/ (as in *decree*). /gri/# is 45 times more common than /kri/# as a token frequency, but only twice as frequent as a type. This minor phonological regularity of English is of interest because Moreton's categorisation-bias experiment yielded a null result, contrary to hypothesis. This result requires scrutiny, because it contrasts with a large body of results indicating that phonotactic knowledge is stochastic. The phonotactic constraint in question combines a triphonemic specification with a less than typical stress template, hence it is similar in complexity to the hypothetical NO-plus-stress constraint.

The relative learnability of each of the four target patterns was calculated as follows: 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 ranked, with individuals viewed as agreeing in their grammars if they agreed in the ranking. If all 20 individuals have the same ranking, then the pattern in question is obviously learnable with estimated $p < .05$ of any error. Any disagreement in ranking means that agreement

TABLE 1
Partial results for learning of NO patterns. Absolute counts in an 800 word vocabulary for 4 individuals

Individual	*/n.t/*	*/n.s/*	*/n.f/*	*/m.f/*	*/n.f/*
1	14	6	3	1	0
2	12	6	4	0	0
3	9	9	1	1	0
4	11	12	0	0	0

among 20 individuals on that ranking cannot be assured, given the vocabulary size.

RESULTS

Stress regularity (1) is perfectly learned for all five vocabulary sizes. In even the smallest vocabulary size examined, the 100 pattern is more common than the 010 pattern in all 20 individual vocabularies. This highly robust bit of phonology has no phonemic specifications, and so all trisyllabic words can contribute to the size of the training set.

Example data on learning of the NO pattern (2) is given in Table 1. This table shows absolute counts for the five heterosyllabic NO clusters, over an 800 word vocabulary, for 4 of the 20 individual vocabularies. Individual no. 1 has five distinct counts which are in the same order as counts in the community vocabulary. Hence, this individual has learned the full ranking. For individuals no. 2 and no. 3, there are ties between cluster counts. Since the corresponding counts in the community vocabulary differ, individuals no. 2 and no. 3 show incomplete learning. Individual no. 4 presents a particularly deviant case. There is a three way tie amongst the less frequent clusters, and there are more words with /n.s/ than with /n.t/ (reversing the correct rank ordering).

Table 2 provides an aggregate measure of the extent to which individuals' vocabularies show the correct ranking, for all five vocabulary

TABLE 2
Mean Spearman rank correlations for NO clusters and NO clusters in strong-initial trisyllables

Vocabulary size	*NO Clusters*	*NO clusters in strong-initial trisyllables*
400	0.80	0.21
800	0.88	0.45
1600	0.98	0.61
3200	1.00	0.68
6400	1.00	0.93

sizes examined. It compares the learnability of NO rankings (2) with the learnability of the NO ranking as confined to the initial trochee of trisyllabic words (3). The aggregate measure of learnability is the mean of the Spearman rank correlation between the phonotactics of individual vocabularies and the phonotactics of the full community vocabulary. This statistic reaches a maximum value of 1.0 when all individuals learn the same ranking as the ranking in the full vocabulary.

Overall, the NO pattern displays a rather good degree of learnability (at 0.8) for even the 400 word vocabulary—indeed the examples in Table 1 are not representative, but were selected to illustrate bad as well as good outcomes. It reaches 1.0 at 3200 words and remains there for the next larger vocabulary. The learnability of the hypothetical NO-stress 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 not all adults achieve, this regularity is not viable as a shared constraint in the phonology of the community.

Turning now to the /gri/#—/kri/# contrast (4), the only relevant monomorphemic words listed in CELEX are *agree, degree, filigree, pedigree* and *scree, decree*. The complete CELEX also contains four complex forms with embedded *agree* or *degree*. In the simulation, only 90% of individuals with a 6400 word vocabulary had learned that /gri/# is better than /kri/#. This outcome occurred because some individuals acquired *scree* and *decree*, but not *pedigree* and *filigree.* In this particular simulation trial, the learnability was actually worse for a 6400 word vocabulary than for a 3200 vocabulary (for which 95% of individuals learned the generalisation). However, even at 6400 words, only 10% of individuals had all four /gri/ words in their vocabulary. Overall, the phonotactic generalisation hangs by the thread of a just a few distinct types. As discussed in Bybee (2001), the experimental and historical evidence suggests that small numbers of types, no matter how individually frequent each one may be, do not tend to project phonological generalisations. Moreton's failure to find a difference between word-final stressed /gri/ and /kri/ in speech perception thus suggests that they are not the subject of phonotactic constraints as such. Instead, their well-formedness is based on the well-formedness of their subparts as exhibited in more diverse positions.

DISCUSSION

The calculations just presented bear out the claim that phonological regularities can be uniformly learned by individuals with different vocabularies. A local regularity mentioning specific phonemes—a five way ranking of NO clusters—can be uniformly learned from 3200 words or less. This is 28% of the community vocabulary, a vocabulary consistent with the suggestion that phonotactics are learned by late childhood or early adolescence. A stress constraint involving a large temporal scale but no specific phonemes can be learned even more readily. The calculations also show that two complex patterns involving a combination of particular phoneme sequences with a specfic stress template are not reliably learnable even from an adult-level vocabulary. Thus, such constraints are not robust enough to use in bottom-up phonological encoding. The relationship of statistical reliability to complexity and vocabulary size appears to be in the right range to correspond to the positive and negative experimental results cited above.

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 Hay et al. (in press) in fact presupposed without argument somewhat finer grained descriptions than used here. Noise in the experimental data makes 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 follow-up work in which individual variation in well-formedness judgements and speech perception is related to assessment of individual vocabularies.

REFERENCES

Atkinson-King, K. (1973). Children's acquisition of phonological stress contrasts. *UCLA Working Papers in Phonetics*, No. 25.

Barton, D. (1980). Phonemic perception in children. In G.H. Yeni-Komshian, J.F. Kavanagh, & C.A. Ferguson (Eds.) *Child phonology: Volume 2, perception* (pp. 97–116). New York: Academic Press.

Baayen, R.H., Piepenbrock, R., & Gulikens, I. (1995). *The CELEX lexical database (release 2) CD-rom.* Linguistic Data Consortium, University of Pennsylvania, Philadelphia PA.

Bybee, J. (2001). *Phonology and language use.* Cambridge: Cambridge University Press.

Chevrot, J.-P., Beaud, L., & Varga, R. (in press). Developmental data on a French sociolinguistic variable: the word-final post-consonantal /R/. *Language Variation and Change*, *12*.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*, 218–236.

Eguchi, S., & Hirsh, I. (1969). Development of speech sounds in children. *Acta Octo-Laryngologica,* Suppl. 257.

Hay, J.B., Pierrehumbert, J., & Beckman, M.E. (in press). Speech perception, well-formedness, and the statistics of the lexicon. In R. Ogden, J. Local, & R. Temple (Eds.), *Papers in laboratory phonology VI.* Cambridge: Cambridge University Press.

Hayes, B. (1994). *Metrical stress theory: Principles and case studies.* Chicago: University of Chicago Press.

Johansson, C. (1997). *A view from language; Growth of language in individuals and populations.* Ph.D. dissertation, University of Lund.

Jusczyk, P. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kent, R., & Forner, L. (1980). Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics*, *8*, 157–168.

Lee, S., Potamianos, A., & Narayan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, *105*, 1455–1468.

McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

Moreton, E. (1997). Phonotactic rules in speech perception. Abstract 2aSC4, 134th Meeting of the Acoustical Society of America, San Diego, CA, Dec. 1–5.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

Norris, D.G., McQueen, J.M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–325.

Raimy, E., & Vogel, I. (2000). Compound and phrasal stress: A case of late acquisition. Paper delivered at the Annual Meeting of the Linguistic Society of America, Chicago, Jan. 6-9.

Vitevich, M., & Luce, P. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*, 325–329.