

PROCEEDINGS OF THE TWENTY-NINTH ANNUAL MEETING OF THE

**BERKELEY LINGUISTICS SOCIETY**

February 14-17, 2003

**GENERAL SESSION**

and

**PARASESSION on PHONETIC SOURCES  
OF PHONOLOGICAL PATTERNS:  
SYNCHRONIC AND DIACHRONIC EXPLANATIONS**

Edited by

Pawel M. Nowak  
Corey Yoquelet  
David Mortensen

439904

**Berkeley Linguistics Society  
Berkeley, California, USA**

## Accepting Markednesslessness: How non-phonological symbolic computation shapes trends in attested phonological systems

CHARLES REISS  
*Concordia University*

### 1. Introduction

This paper makes some programmatic suggestions concerning how properties of general auditory processing provide insight into the kinds of sound changes that recur in the transmission of language, thus giving rise to the kinds of recurring patterns seen in the phonological systems of the world's languages. The position developed here must be clearly distinguished from any attempt to build phonetics into phonology, or to blur the distinction between various domains—in fact, one subgoal of this work is to develop coherent and useful definitions of various domains of inquiry, for example, synchronic linguistics vs. diachronic linguistics (two fields of study, one of which examines the human language faculty) and auditory perception vs. phonology (both of which examine faculties of the mind, one that interfaces with the language faculty, and one that is a component of the language faculty).

I hope to show that rejecting substantive 'markedness' (cf. Hale & Reiss 2000ab) as a crucial part of phonological theorizing does not force us into theoretical nihilism. There is still plenty of work to do in 'psychological phonology', and there is still plenty to do in related fields, even if we accept 'markednesslessness'.

Drawing on recent work in auditory perception in the auditory scene analysis framework, I hope to support the research program founded by John Ohala which aims to explain phonological patterns by appeal to extragrammatical phenomena, as alluded to in this recent quotation:

So, the phonetic primitives invoked in the modeling of these [diachronic] processes make no pretenses of being psychological. The attempts by

those who are interested in psychological phonological grammars and in finding ways to represent phonological processes ... in phonetically natural ways have been abysmal failures ... One possible solution to this is not to put more phonetic sophistication into psychological grammars but rather to abandon phonetic naturalness as a necessary feature of them.

Ohala 2002:685

This approach is to be contrasted with the currently dominant trend within linguistics of building more and more phonetics into phonology, for example, in the form of phonetically grounded grammatical constraints.

## 2. Background assumptions and terminology

Following Keating (1988) I adopt the view that phonology is most usefully defined to consist of a particular kind of symbolic computation. The input and output of the phonology are representations described in terms of a single symbolic alphabet. The output of the phonology is *transduced* to actual articulatory and perceptual events via a series of transducers, some of which are also cognitive, but by definition non-phonological. This modular approach to explanation is consistent with standard views within cognitive science:

In using the word 'representations', we are implying the existence of a two-part system: one part forms the representations and another uses them to do such things as calculate ...

Albert Bregman, *Auditory Scene Analysis* (1990:3)

It is crucial to note that cognition builds and manipulates EQUIVALENCE CLASSES—it is not the case that every physically distinguishable contrast is computationally/psychologically relevant (cf. Pylyshyn 1984, and even Sapir 1933).

In order to contrast the focus of this paper with what I take to be useful markednessless phonological theorizing, I will mention two examples of the latter. Work by Sharon Inkelas (e.g., 1996) on the form of featural representation in the lexicon, including issues such as the justification for underspecification constitutes valid, interesting phonological work. My own work attempting to demonstrate the necessity of the use of quantifiers in phonological computation and the implications of this necessity, viz., that feature geometry is *not* a property of phonological representations is another example of work that falls squarely within the tradition of phonological reasoning in a generative framework. In the next section we turn from phonology to a discussion of auditory perception.

## 3. Auditory Scene Analysis

The perceptual world is one of events with clearly defined beginnings and endings ... An event becomes defined by its temporal boundary. But this impression is not due to the structure of the acoustic wave; the beginning and ending often are not physically marked by actual silent intervals.

Handel, 1989

Auditory scene analysis is a framework for studying auditory perception developed by Albert Bregman and his collaborators. I think Bregman would be the first to admit that work in the field is still in its infancy. However, it has now become possible to ask questions concerning the nature of auditory perception that approach the sophistication of question in domains such as visual perception.

Auditory scene analysis can be broken down into two main components. One problem, given the fact that sounds waves from various sources are combined into a single wave that reaches the eardrum, is that of simultaneous integration and segregation. In other words, the auditory system integrates into a single representation parts of the sound spectrum reaching the ear within a temporal window that 'go together', according to the properties of the system (nonveridically, in the case of an illusion). The process of assigning parts of the spectrum to different perceptual sources is called spectral segregation.

The other main component of auditory scene analysis is sequential integration—acoustic events occurring separated in time may be integrated into a single auditory stream. Examples of streams include a sequence of footsteps or the continuous sound of falling rain.

### 3.1. A Simple Grammar for auditory scene analysis

Building on the work of Y. Nakajima and T. Sasaki, Nakajima (1996) proposes that auditory scene analysis can best be understood in terms of a grammatical system that recognizes elements smaller than the auditory events (a footstep, a musical note, etc.) that make up auditory streams. These elements, auditory subevents, are classified into at least four fundamental types.

#### (1) Types of auditory subevents

- Onset (denoted by <): A steep rise of sound intensity within a certain frequency range (e.g., a critical band) can be a clue of an onset.
- Termination (denoted by >): A steep fall of sound intensity within a certain frequency range can be a clue of a termination.

- Filling (denoted by =): A piece of sound energy extending for a certain duration without any sudden change of frequency range can be a clue of a filling.
- Silence (denoted by /): If the sound energy across a certain frequency range and a certain duration is very thin despite some amount of sound energy in the preceding part, this makes a clue of a silence.

Onsets and terminations will be referred to below as boundaries.

As these definitions make clear, what is of interest to us are the cognitive representations transduced from the physical signals. So, for example, the auditory grammar subevent 'silence' can be transduced in a context of *acoustic* energy, for example, where the energy is not part of the same stream as the perceptual 'silence'. It is sometimes difficult in discussion to be consistent in distinguishing the cognitive subevents from their acoustic correlates. However, use of the symbols mentioned above will be helpful: the symbol '/' is a constructed representation that may be transduced during a period of acoustic energy, for example.

According to Nakajima, most auditory events in our everyday life seem to take one of the three modes indicated below.

(2) Auditory events:

- An onset followed by a silence (e.g., a clap sound): ( $</$ ),
- An onset, a filling and a termination followed by a silence (e.g., a cat's meow): ( $<=>/$ ),
- An onset and a filling followed by the onset of another auditory event (e.g., a note in a melody played legato): ( $<=><$ ).

Following Nakajima, the auditory events we find are represented by the portions within parentheses. The subevents, following the right parenthesis represent the beginning of the next auditory event or a silence.

The primitive auditory subevents are thus grouped into events which are the immediate constituents of auditory streams. The events are combined into streams according to a small set of principles.

(3) Definition of auditory stream:

- An auditory stream is a linear string of auditory events and silences in temporal order.
- An auditory stream begins with an onset and ends with a silence.
- A silence is not followed immediately by another silence.

To summarize the model developed thus far, an auditory scene is composed of auditory streams that are composed of auditory events that are composed of auditory subevents.

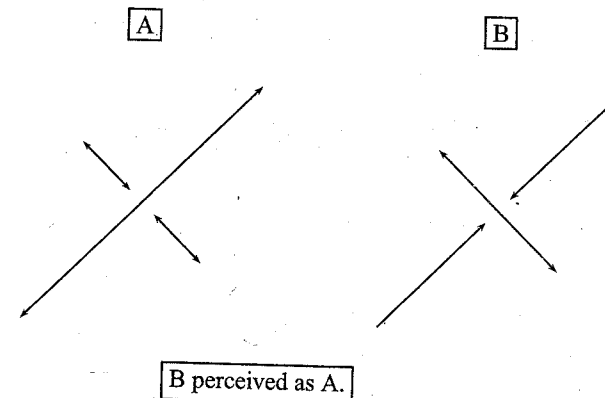
#### 4. Gap transfer illusion

In this section we discuss an auditory illusion, the gap transfer illusion, discovered by Nakajima and Takayuki Sasaki that suggests that "When it is difficult to interpret all the given clues of auditory subevents grammatically, the auditory system inserts new subevents, interprets the same clues twice or more, or suppresses some clues." It is crucial to keep in mind that the subevents are discrete symbols over which auditory scene analysis computes. The four primitives are manipulated by the computational system in the construction of the auditory scene. Gestalt principles such as proximity and similarity can explain some of the manipulation of subevents by cognition.

Our discussion of the gap transfer illusion will necessarily be rather informal. In the diagrams below, arrowheads are to be equated with the boundary symbols,  $<$  and  $>$ , introduced above. Lines correspond to fill ( $=$ ). The x-axis corresponds to time and the y-axis to frequency.

The gap transfer illusion arises when a stimulus such as (4B) is perceived as (4A).

(4) Gap transfer illusion



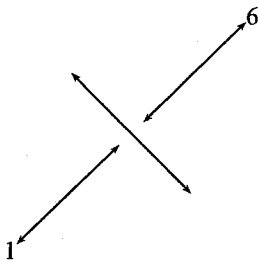
That is, the gap in the acoustic rising glide is transferred to the falling glide.

There are actually several things going on here, some of which are strictly speaking not part of the gap transfer effect. It will be useful to discuss them, however, to illustrate the working of the auditory grammar that Nakajima and his colleagues propose. First, note that the stimulus in (4B) does not begin with an onset,  $<$ , or end with a termination,  $>$ . This is because the rise and fall of the sound intensity in these locations were not 'steep enough to give such clues', according to Nakajima. A filling at the beginning of an auditory stream or a filling immediately followed by a silence is ungrammatical. That is, such contexts cannot be derived

from the above-mentioned grammatical rules. Apparently, the solution for the auditory system is to insert an onset (in accordance with (3b)) and a termination (in accordance with (2b)).

The insertions demanded by these auditory rules would give something like (5). (Note that I am fudging the issue of what exactly (5) is. It is not exactly a representation of the stimulus and it is not the final percept. Perhaps it is best thought of as an intermediate auditory perception.)

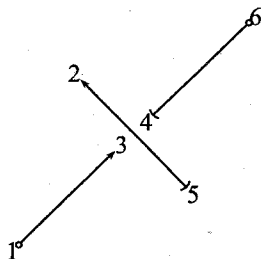
(5) Boundary insertion



The relevant onset and termination are inserted by the auditory grammar at the points labelled '1' and '6'.

We are now ready to describe the Gap Transfer Illusion *via* the Gestalt Principle of Proximity. In order to do this, I have replaced some of the arrowhead onset and termination markers with other symbols. The recoupling or reassociation of these boundary markers can be deduced by matching pairs of symbols. The discussion will also be facilitated by the numerals, corresponding to the physical temporal sequencing, that I have assigned to the boundary markers.

(6) Boundary reassociation



In the signal, or, more precisely, the intermediate representation of (5), boundaries 1 and 3 belong together, as the subscript  $\alpha$  denotes below in (7). Boundaries 2 and

5, with subscript  $b$  also belong together. Finally, 4 and 6, with subscript  $c$ , belong together.

(7) SIGNAL (plus perceptual insertion of 1 and 6)

$\langle a_1 \quad \quad \quad \rangle \quad \langle b_2 \quad \quad \quad \rangle \quad \langle c_4 \quad \quad \quad \rangle \quad \langle b_5 \quad \quad \quad \rangle \quad \langle c_6 \quad \quad \quad \rangle$

However, the percept is very different. The temporally close onset-termination pair (2,3) have been coupled, as indicated by the shared arrowheads in (6) and by the subscript  $\beta$  that they share in (7). Similarly, the pair (4,5) share the curved boundary markers in (6) and the subscript  $\gamma$  in (7), denoting the fact that they have been perceptually coupled. This leaves the pair (1,6), with open circles for boundary markers in (6) and subscript  $\alpha$  in (7).

(8) PERCEPT (after reassociation of boundaries)

$\langle \alpha_1 \quad \quad \quad \rangle \quad \langle \beta_2 \quad \quad \quad \rangle \quad \langle \gamma_3 \quad \quad \quad \rangle \quad \langle \gamma_4 \quad \quad \quad \rangle \quad \langle \beta_5 \quad \quad \quad \rangle \quad \langle \alpha_6 \quad \quad \quad \rangle$

The percept resulting from insertion and reassociation of boundaries from an input stimulus like (4B), is thus identical to that resulting from an input stimulus like (4A).

Before moving on to speech perception, let's make two observations. First, the percepts of onsets and terminations, as well as their mutual associations, are *constructed* in the process of auditory stream analysis.

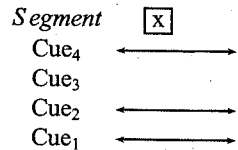
Second, it is suggestive that the effect of reassociation in the gap transfer illusion is to provide the auditory event with an immediate constituent analysis. In the final percept, as denoted in (8), constituent  $\alpha$  contains constituents  $\beta$  and  $\gamma$ , and  $\beta$  precedes  $\gamma$ , but there are no interlocking constituents. On a highly speculative note, I suggest that this is the defining characteristic of auditory streams, as well as of linguistic representations. In fact, the perception of speech as a sequence of discrete segments may be a reflection of this property of auditory streams—in the streaming of the diverse components of the speech signal only precedence and containment relations among cues are licit.

## 5. Application to speech perception

In this section, we treat speech signals in isolation as auditory scenes composed of various auditory streams that correspond to the acoustic cues generated by a human vocal tract in speech. In other words, acoustic parameters such as the value of the first formant or the presence of high frequency broadband noise, each constitute, by hypothesis, an auditory stream. These streams, of course, can be further analyzed into their component events and subevents.

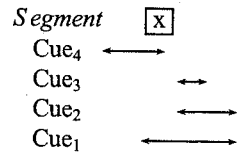
We can imagine an idealized representation of the relationship between the acoustic cues that are transduced to phonological featural representations with perfect temporal alignment for all the cues. In the example in (9) Cue<sub>3</sub> is absent from segment *x*.

(9) Idealized segment



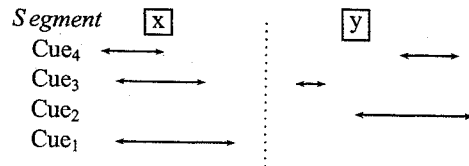
However, it is well known that from either an articulatory or an acoustic perspective, temporal relations of gestures or perceptual cues are much less orderly, as shown in (10), where Cue<sub>1</sub> extends over Cue<sub>3</sub> completely, and Cue<sub>1</sub> partially overlaps with Cue<sub>4</sub>, and the latter completely precedes Cue<sub>1</sub>.

(10) Acoustic cue / articulatory gesture alignment



Even granting that cues do not all line up perfectly in duration, we could imagine an idealized representation of the temporal sequencing of cues in a sequence of segments, with a neat division between the cues belonging to two segments *x* and *y*, as in (11):

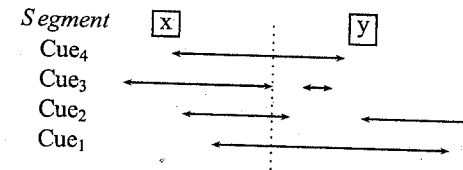
(11) Idealized alignment of cues / gestures



However, once again, the acoustic and articulatory reality is much messier. The alignment relations in (12) are much more realistic.

(12) Physical alignment of cues / gestures

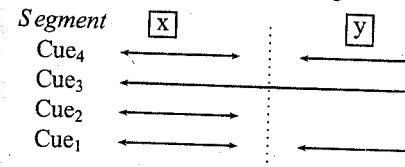
*Accepting Markednesslessness*



In other words, cues that end up being interpreted as belonging to a give segment may partially or even fully overlap temporally with the cues of another segment.

Despite the complex temporal relation among cues both within and across segment boundaries, I would like to suggest that the equivalence classes generated in the process of speech perception lead to a representation more like (13), where I have equated (simplistically) cues with features. In other words, cues are transduced to feature bundles (segments) which are discrete categorical symbols.

(13) Claim: (Quasi-)Phonological representation



We can now proceed to a discussion of how auditory scene analysis and Nakajima's grammar of auditory events can give insight into the explanation for the recurring patterns observed in phonological systems.

### 5.1. Why are long cues subject to change?

In various works John Ohala and those influenced by him (e.g., Blevins 2003, Blevins and Garrett 2000) have noted that phonological features corresponding to phonetic cues with long durations are more likely to undergo common sound changes such as assimilation, dissimilation and metathesis. Examples cited by Blevins & Garrett include rhoticity (the acoustic correlate of which is lowered F3), palatalization (raised F2), and rounding (lowering of all formants). I would like to make the simple suggestion that the explanation for this behavior is that the acoustic signals corresponding to streams for long duration cues will overlap with more subevents belonging to other streams than the acoustic signals of streams for short duration cues. Thus, there are more possibilities that the boundaries of long duration cues be reassociated by a perceiver with acoustic subevents belonging to other streams in the target language.

Consider first how the Proximity Principle could be invoked to account for assimilation. In (14.i) we represent abstractly two cues/features *p* and *q*, that are

attributes of separate segments in the target language. The vertical line is meant to denote a segment boundary.

Now suppose that (14.ii) represented the temporal relations of the boundaries (onsets and terminations) of these cues as produced by speakers of the target language. Cue  $p$  corresponds to constituent  $a$  and cue  $q$  corresponds to constituent  $b$ , and there is a partial overlap of  $p$  and  $q$ . Under appropriate conditions (to be determined by future researchers), the proximity of onset  $<_2$  and termination  $>_3$  leads to association of these two subevents by the learner/perceiver. If we assume that boundaries  $<_1$  and  $>_4$  are also coupled, we get a model of the learner's auditory percept that has the structure of (14.iii), where constituent  $\alpha$  contains constituent  $\beta$ . Step (14.iv) shows how this percept is transduced to a grammatical representation.

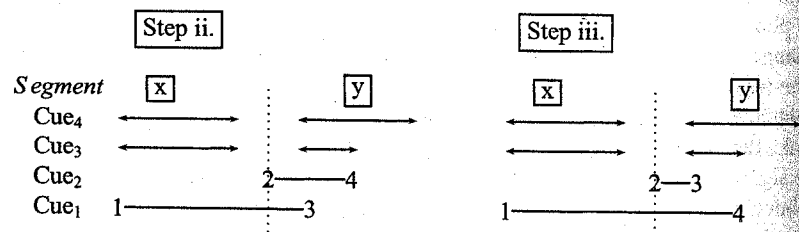
#### (14) Assimilation (via Proximity-driven reassociation)

- i. FEATURAL REPRESENTATION IN TARGET LANGUAGE  $[p] | [q]$
- ii. CUE PATTERN IN TARGET OUTPUT  $<_{a1} \quad <_{b2} \quad >_{a3} \quad >_{b4}$
- iii. LEARNER'S AUDITORY PERCEPT  $<_{\alpha 1} \quad <_{\beta 2} \quad >_{\beta 3} \quad >_{\alpha 4}$
- iv. LEARNER'S GRAMMATICAL (FEATURAL) PARSE  $[p] | [p, q]$

In this grammatical representation, the first segment is still associated with the cue/feature  $p$ , and the second with  $q$ , but the second is also associated with  $p$ . If the learner stores the representation thus, diachronic assimilation has occurred.

Steps (ii) and (iii) can also be represented as follows, with  $p$  corresponding to cue 1, and  $q$  to cue 2.

#### (15) Temporal relations of cues

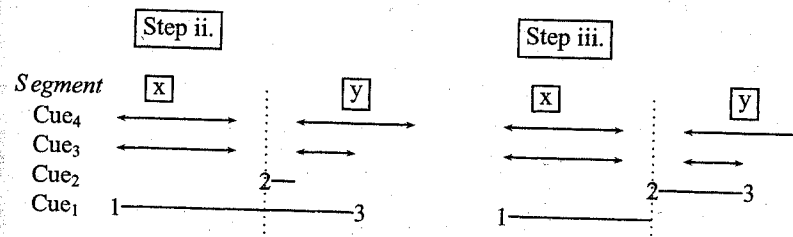


Parallel to the case of the gap transfer illusion, the association of 2 and 3 allows 1 to associate with 4.

It is not difficult to extend this form of reasoning to derive dissimilation of a long duration cue. In the following diagrams, the target language has a segment associated with cue  $p$  followed by one associated with  $p$  and  $q$ . The auditory event corresponding to  $q$  either has no termination, or else the termination is masked by other acoustic properties of the stimulus. The dissimilatory change is the loss of association of  $p$  with the second segment.

#### (16) Dissimilation (via Proximity-driven reassociation)

- i. FEATURAL REPRESENTATION IN TARGET LANGUAGE  $[p] | [p, q]$
- ii. CUE PATTERN IN TARGET OUTPUT  $<_{a1} \quad <_{b2} \quad >_{a3}$
- iii. LEARNER'S AUDITORY PERCEPT  $<_{\alpha 1} \quad <_{\beta 2} \quad >_{\beta 3}$
- iv. LEARNER'S GRAMMATICAL (FEATURAL) PARSE  $[\alpha] | [\beta]$



The association of 2 and 3 leads to the creation from 1 of an event of type (2c), 'an onset and a filling followed by the onset of another auditory event', so Cue<sub>1</sub> does not extend through the second segment in the learner's parse.

#### 6. The CRUM-my Conclusion

Obviously much more work is needed to make the suggestions here more concrete. The application of the Principle of Proximity and its interaction with other Gestalt principles is poorly understood, and it will be necessary to look at real examples of assimilation and dissimilation to see if Nakajima's auditory grammar will work for a signal as complex as speech. However, the notion of unifying aspects of phonology with aspects of speech perception and ultimately with aspects of auditory perception is appealing and warrants further research. Note that this unification does not entail a rejection of phonology as a domain of inquiry—aspects of phonological systems (and phonological typology) that can be explained by reference to audition do not have to be explained by reference to the faculty of phonological computation. This leads to a more elegant theory of phonology.



I adopt the position of classical cognitive science that cognition can best be understood in terms of representational structures composed of discrete, categorical primitives and computational procedures that operate on those structures: this view is dubbed the Computational-Representational Understanding of Mind (CRUM) by Thagard (1996). Generative phonological theory is an example of a CRUM-my theory. (That is a good thing.)

The thesis of this paper is that the recurring patterns (assimilations, dissimilations, etc.), that we find in phonological systems are due to a CRUM-my module of mind. However, those patterns are the result of the CRUM-my nature of auditory scene analysis applied to speech during acquisition, and not the computational properties of the phonological component. This position is consistent with the diachronic, nonfunctionalist approach to sound change and phonological typology represented by the work of Ohala, Blevins & Garrett and Hale.

I conclude then, that it is not necessary to build the details of phonetic substance into the phonological component by encoding 'markedness' based on typological tendencies into grammar. There is plenty of real phonology to be done, and there is plenty of work in fields such as auditory perception and speech perception. In other words, a markednesslessnessless phonology is not needed to combat nihilism.

One final observation should be made. If we accept the analysis of the gap transfer illusion presented by Nakajima and his colleagues, then we accept the possibility that the auditory system manipulates entities like terminus markers in a way that is highly abstracted from their physical substance, however that may be defined. It is no surprise therefore that phonology, which in turn is several steps removed from auditory processing should also show evidence of substance-free computation.

## References

- Blevins, Juliette. To appear. *Evolutionary Phonology*. Cambridge University Press.
- Blevins, Juliette and Andrew Garrett. To appear. The evolution of metathesis. In Bruce Hayes, Robert Kirchner, and Donca Steriade (eds.). *Phonetically driven phonology*. Cambridge University Press.
- Blevins, Juliette and Andrew Garrett. 1998. The origins of consonant-vowel metathesis. *Language* 74: 508-56.
- Bregman, A. 1990. *Auditory Scene Analysis*. Cambridge, MA: MIT Press.
- Hale, M. To appear. *Theory and Method in Historical Linguistics*. Blackwell.
- Handel, S. 1989. *Listening*. Cambridge, MA: MIT Press.
- Inkelas, Sharon. 1996. Archiphonemic Underspecification. Ms. UC Berkeley.
- Nakajima, Y. 1996. A simple grammar for auditory organization: streams, events and subevents. *Approaches to Auditory Organization" XXVI International Congress of Psychology*, Montreal, Canada
- Nakajima, Y., Sasaki, T., Kanafuka, K., Miyamoto, A., Remijn, G., and ten Hoopen, G. (2000). Illusory recouplings of onsets and terminations of glide tone components. *Perception & psychophysics*, 62, 1413- 1425.
- Pylyshyn, Z. 1984. *Computation and cognition: toward a foundation for cognitive science*. MIT Press, Cambridge, MA.
- Reiss, Charles. 2003. Quantification in phonological rules. *The Linguistic Review* 20: *Special Issue on Typology in Phonology*, Spring 2003.
- Sapir, E. 1933. La Réalité Psychologique des Phonèmes. *Journal de Psychologie Normale et Pathologique* (Paris), 30: 247-265.