

Modelling phonological category learning

by Paul Boersma, 10 August 2010

10.2.1. What is category learning?

The term “phonological categories” refers to the discrete elements that make up a phonological representation, i.e. elements of its temporal organization (e.g. foot, syllable, mora, segment, or autosegment) and elements of its internal content (e.g. phonemes such as /p/ and /n/, or feature values such as [+nasal] and the high tone H). From the mere fact that the term “phonological category learning” appears in the title of this contribution, the reader can already infer that the editors of this Handbook, and/or the present author, assume that at least some of these phonological categories can be learned. This assumption is opposite to the assumption held by most generative phonologists, which is that all phonological categories are innately given to the human infant. Thus, Chomsky & Halle (1968: 4) state that “phonetic features” belong to the “substantive universals”, which are a subgroup of “linguistic universals”, which are “available to the child... as an a priori, innate endowment”. Likewise, Prince & Smolensky (1993 [2004: 2–3]) state that “Universal Grammar”, i.e. the innate language endowment, “consists largely of a set of constraints on representational well-formedness”; in their examples, such innate constraints often refer to substantive phonological elements, which therefore have to be innate *a fortiori*.

As has been pointed out for syntax and semantics by Braine (1992), Slobin (1997: 289–296), and Tomasello (2003: 183–185), the generative assumption of innate categories comes with a learnability problem, namely the *linking problem*. In the phonology and phonetics domain this means that even if all phonological categories were innately given, the language-acquiring child would still have to connect at least some of these innate categories to auditory events available in the incoming speech data, and this is a problem because the mappings between some phonological categories and auditory events vary crosslinguistically and cannot therefore be innate. After all, the hypothesis of innate categories presupposes the universal existence of e.g. the phoneme category /u/ (or of the feature values [+back], [+high] and [+round]), but since a phonological element representable as /u/ (or as the feature bundle [+back, +high, +round]) is typically pronounced slightly differently in every language, the mapping between this phonological category and auditory events must be language-specific and cannot therefore be innately given.

The reasoning in the previous paragraph may not convince many generative phonologists. After all, a generative phonologist could object that an innate category /u/ could correspond to a *region* of auditory events, e.g. a cloud of F1–F2 pairs, and that different languages select different parts of this cloud. This objection fails if one realizes that the perceptual *boundary* between e.g. the vowels /u/ and /o/ is also language-specific, i.e. two different languages should be able to both have the innate categories /u/ and /o/, but there will be some sounds that are perceived as /u/ by listeners of one language and as /o/ by listeners of the other language. This abundantly established fact (for direct crosslinguistic comparisons, see Savela 2009

for vowels or Hamann, Boersma & Čavar 2010 for the [f]–[v]–[ʋ] continuum) proves that at least some categories cannot be innately connected to specific sounds.

It might still be possible to hold the innatist viewpoint here and devise a learning algorithm that starts in some default sound-to-category-connection state and subsequently shifts the category boundaries on the basis of incoming speech data, analogously to ideas known from the syntax-semantics interface such as Grimshaw's (1981) innate "canonical structural realizations" or Pinker's (1984, 1989) "bootstrapping" with innate linking rules. To my knowledge, however, no such algorithm has been explicitly proposed in the phonological-phonetic literature. For this reason, I will in the remainder of this contribution indeed assume the *emergentist* viewpoint of category learning, which holds that the language-acquiring child comes without any innate phonological categories and subsequently *creates* her categories on the basis of incoming speech data.

10.2.2. Where do categories emerge?

Assuming, then, that phonological categories emerge in the language-acquiring child, the question is in what location (representation in the brain) these categories emerge. A priori, it would be good to have phonological categories in the phonological lexicon, which is the location in which humans typically have to store enough sound information to make thousands of morphemes pronounceable and perceivable: categories are discrete internal representations of raw continuous data in the outer world and can thus provide a helpful reduction of the required data storage. For this reason, most psycholinguists and phonologists agree that the phonological lexicon is built of discrete categories (for the opposite standpoint of exemplar theory, which is less concerned about lexical economy, see below in §10.2.7).

But is the lexicon the *only* location where these categories exist? There are three sources of evidence that phonological categories exist even outside the lexicon. The first source of evidence comes from psycholinguistic experiments; psycholinguists with quite diverging convictions on the details of phonological comprehension (McClelland & Elman 1986; Samuel 1996; Norris, McQueen & Cutler 2000) can agree that the speech comprehension process goes through a *prelexical representation* that consists of the same kinds of phonemes (and other phonological elements) as the lexical representation; the basic idea is that human beings in the lab can readily identify phonemes in tasks that do not involve access to the lexicon, such as tasks involving short syllables that are not words. The second kind of evidence comes from phonological theory, where it is widely agreed (e.g. Prince & Smolensky 1993) that the speech production process goes through a *phonological surface structure* consisting of discrete phonological elements such as feet, syllables, segments, and features; the basic idea is that especially the larger metrical structures (feet, syllables) cannot be specified in the lexicon (the *underlying form*), because the domain of their assignment is often the phrase rather than the word (i.e. these structures tend to span across word boundaries). The third kind of evidence comes from infant studies, which find that the perception of children of 8 to 10 months of age is already adapting to the phonological categories of their ambient language environment (Kuhl 1991; Polka & Werker 1994; Jusczyk 1997); the basic idea is that although these infants have no words in their lexicons yet, they have already increased the ability to distinguish

between sounds that belong to different phonological categories and decreased the ability to distinguish between sounds that belong to the same phonological category.

When we combine the three sources of evidence, and assume that humans of any age have the same levels of representation, the simplest hypothesis must be that categories emerge in the intermediate level (the prelexical representation or phonological surface structure), and that this happens in the infant's comprehension process, and more specifically in the infant's acquisition of her (prelexical) perception. This is shown in Figure 1.

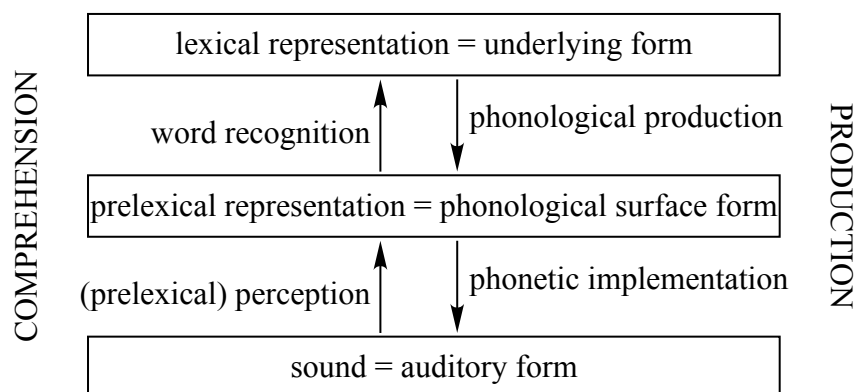


Fig. 1. The simplest model of speech comprehension and production compatible with the evidence from psycholinguistic experiments, phonological theory, and infant studies. Categories emerge in the intermediate level, as a result of the acquisition of perception.

10.2.3. What do categories emerge from?

In Figure 1, the phonetic correlate of a phonological category is auditory. Although this is in line with the acquisitional evidence discussed in 10.2.2 (infants perceive contrasts before they articulate them), the possibility that the phonetic correlate of a phonological category is instead articulatory, especially in production, cannot be ruled out. In linguistics the auditory view is shared by Saussure (1916) with his *image acoustique* and by Jakobson, Fant & Halle (1962) and, from a nativist camp, Anderson & Ewen (1987). Since the phonetic implementation process must also somehow feed into articulation, this view has to entail that articulation happens in the service of audition. Thus, Harris & Lindsey (1995) argue for the primacy of audition from biteblock experiments (Lindblom, Lubker & Gay 1979), in which speakers maintain auditory forms by modifying their articulations (for a slightly different view see Folkins & Zimmerman 1981); in an explicit model of phonological-phonetic production, Boersma (1998) places the articulatory form below the auditory form in Figure 1, arguing that for implementing the phoneme /s/ the auditory correlate of loud high-frequency noise is primary whereas the articulatory correlate of alveolar constriction is secondary, the idea being that in order to articulate a legitimate /s/ you also have to make sure that your lungs contract, your glottis is wide, your velum is up, and your lips are open, everything in service of the production of auditory loud high-frequency noise.

Many authors (Chomsky & Halle 1968, Clements 1985, Browman & Goldstein 1989, Keyser & Stevens 1994), and therefore probably many readers of these lines, do not share the auditory view of phonetic implementation: they assume instead that the

phonetic correlate of phonological categories is articulatory in nature. In Saussure (1916: 98), for instance, Bally & Secheyave found it necessary to include a footnote explaining Saussure's standpoint against the articulatory bias of those days, and Ramus *et al.* (to appear) mention Boersma's model but deviate from it (without argument) by positing in their boxes-and-arrows model the articulatory rather than the auditory form as the direct output of phonetic implementation. Ramus *et al.* do not provide an explicit, let alone computational, account of how production or comprehension could proceed; my prediction is that attempts to devise an explicit account of the production of /s/ would fail in the case of their model. I like to stress here that boxes-and-arrows graphs can be verified or falsified only by explicit, preferably computational modelling, something that very few psycholinguistic accounts presented at LabPhon conferences provide. By contrast, phonological accounts by linguists do tend to be fully explicit (e.g. with ordered rule sets or with ranked constraint sets), and therefore have the desirable level of explicitness. Of course, I do agree with Ramus *et al.*'s point that linguists should address not just what Figure 1 calls "phonological production" (as e.g. Prince & Smolensky 1993 do), but also "word recognition" (as Smolensky 1996 does), "prelexical perception" (modelled explicitly by Boersma 1998 *et seq.*, Pater 2004, Berent *et al.* 2009), and "phonetic implementation" (Boersma 2007, 2009; Boersma & Hamann 2009).

The traditional bias in favour of articulatory correlates in production has been extended to comprehension. The hypothesis of *direct realism* (Fowler 1986; Best 1995), for instance, maintains that listeners directly perceive the speaker's articulatory gestures, and the *motor theory of speech perception* (Lieberman & Mattingly 1985) holds that listeners access their phonological forms only after activating their *own* articulatory gestures. In these two models, then, even the left side of Figure 1 would have to be extended with an articulatory level (either the speaker's or the listener's) between the auditory and surface forms. While such extensions are imaginable, the great majority of explicit models of category creation only consider the lower two levels of Figure 1, and it is those models that I discuss in this contribution.

Another issue relevant to how Figure 1 relates to category creation is whether the arrows on the left and on the right represent separate modules or not. According to Ramus *et al.* (to appear), for instance, the arrows "word recognition" and "phonological production" must be separate, because in foreign-language perception Japanese listeners insert vowels but in their phonology they do not (Polivanov 1931, Dupoux *et al.* 1999, Jacquemot *et al.* 2003). There are two things wrong with this reasoning. First, in Smolensky's (1996) explicit (namely, Optimality-Theoretic) bidirectional model, where word recognition and phonological production employ the same ranked relations, insertion in comprehension corresponds to deletion in production (again, we see a dramatic example of why the common LabPhon practice of translating the results of psycholinguistic experiments to boxes-and-arrows plots must fail without an explicit model of what the boxes and arrows mean). Second, the psycholinguistic evidence shows that Japanese perceptual vowel insertion takes place in the module of "prelexical perception", i.e. at a different *level* (not a different *direction*) than phonological production (for an explicit Optimality-Theoretic account of such cases, see Boersma 2009 for Japanese and Boersma & Hamann 2009 for Korean); on the right side of Figure 1, this perceptual capability of inserting vowels

corresponds to the capability of Japanese speakers to delete vowels in “phonetic implementation”, which is an uncontroversial aspect of Japanese pronunciation (Akamatsu 1997). There thus does not seem to be any strong evidence against the bidirectionality proposed by Smolensky (1996) for the top two arrows and by Boersma (2007) for the bottom two arrows in Figure 1; if this bidirectionality is true, categories created on the basis of correct prelexical perception can be employed immediately in phonetic implementation, with correct auditory targets (i.e. potentially hampered only by articulatory effort).

The last issue with Figure 1 is whether the two arrows at the left or right represent sequential modules or not. Interactive (top-down) influences of the lexicon on phonological categorization in comprehension would make at least the “word recognition” arrow bidirectional (for an explicit model see McClelland & Elman 1986), and interactive (bottom-up) influences of phonetic considerations such as articulatory effort and the quality of auditory cues on phonological production would make at least the “phonetic implementation” arrow bidirectional (for an explicit model see Boersma 2007, 2008).

10.2.4. How do categories emerge?

If categories emerge in the phonological surface form (the intermediate level in Figure 1), then one or more other representations must play a role in this process. The simplest computer simulations of phonological category creation (e.g. Guenther & Gjaja 1996; Boersma, Escudero, Hayes 2003) indeed assume that the discrete phonological categories emerge in the surface form from continuous auditory representations such as formants, pitches, durations, noises, silences, and their combinations and sequences (which are in the lowest level in Figure 1). That this modelled procedure is realistic has been confirmed in artificial-language-learning studies such as that by Maye & Gerken (2000) and Maye, Werker & Gerken (2002). We can conclude that bottom-up processing in speech comprehension plays a major role in category creation.

What also might play a role in category creation are all the representations *above* the phonological surface form, not only the underlying form in Figure 1, but perhaps also the syntactic and semantic representations, which must be located even further above. Whether the lexicon plays an active role in determining a perceived category in online *comprehension* is a matter of vigorous debate (e.g. Norris, McQueen & Cutler 2000 versus Samuel 1996), but it is more widely accepted that the lexicon (perhaps via higher-level representations) can act afterwards as a correcting supervisor telling the listener what she should have perceived, because this kind of top-down processing in perceptual *learning* has been observed in the lab (Eisner 2006; Eisner & McQueen 2006). Many explicit models of perceptual learning, e.g. the TRACE model by McClelland & Elman (1986) and an Optimality-Theoretic model by Boersma (1998), therefore include such a supervising mechanism. However, such supervision can only occur once the categories exist, and it is possible that top-down processing plays no role whatsoever in the *creation* of categories.

10.2.5. Requirements for a model of category emergence

Despite the fact that eight-months-olds can profit little from higher representations when creating their first phonological categories, the ultimate comprehensive model of category creation will probably have to be embedded in a larger model that can handle not only the creation of phonological categories and the acquisition of the connections of those categories to auditory cues, but also the acquisition of their connections to higher representations. Such a larger model therefore should not just do category learning but also exhibit many “effects” known from the literature on psycholinguistics, phonological theory and infant studies, such as perhaps the Ganong effect (Ganong 1980), the McGurk effect (McGurk & McDonald 1976), the prototype effect in best-token experiments (Johnson, Flemming & Wright 1993), the perceptual magnet effect (Kuhl 1991), the relation between phonological activity and frequency (what phonologists call “markedness”), auditory dispersion (Liljencrants & Lindblom 1972), licensing by cue (Steriade 2001), and so on. After all, all these phenomena appear in the same language-processing brain, and we should not have to create a separate model for every phenomenon that we observe. Hence, all these phenomena should ultimately be viewed in relation to each other.

If the ultimate larger model is as emergentist as the category creation model has to be, this causes a problem for the hypotheses of direct realism and motor theory discussed in §2, because the fact that infants can categorize before they can speak may require those models to assume an innate connection between sound and articulation. In the following I therefore assume the simpler model of Figure 1, and also assume that all parts of it are emergent.

10.2.6. Existing models of emergence (but not of categories)

Some comprehensive emergentist models do exist already. The neurobiologically inspired TRACE model (McClelland & Elman 1986) considers the three levels of Figure 1 and derives several effects, including the Ganong effect. The present author’s linguistically inspired Optimality-Theoretic model of bidirectional parallel multi-level constraint competition (for an overview, see Boersma to appear) brings together the seven effects mentioned in §10.2.5 under one umbrella: the Ganong effect results from parallel multi-level evaluation, the McGurk effect from Optimality-Theoretic interactions between auditory and visual inputs, the prototype effect and auditory dispersion from the idea that constraint rankings optimized for perception are reused in production, and markedness effects and licensing by cue from a bidirectional multi-level learning algorithm. It has to be remarked here that that model does *not* handle category creation, nor its developmental precursor, the perceptual magnet effect.

10.2.7. Existing models of category creation (but not of phonology)

There exist several models that can handle category creation, although these have rarely been applied to the learning of phonological categories, let alone been embedded within a larger model of language processing. Adaptive Resonance Theory (Grossberg 1976, 1980, 1987; Carpenter & Grossberg 1991) proposes that a new category is created at a certain level of representation (e.g. the phonological surface form in Figure 1) as soon as the brain detects a mismatch between bottom-up

information to that level (e.g. from the auditory form in Figure 1) and top-down expectations (e.g. from the lexical representation in Figure 1). It would be interesting to see how these complicated models perform within a large linguistic model.

Connectionist models also hold a promise of providing mechanisms for category creation. McClelland & Rumelhart (1986) show that if categories are not represented as unitary symbols, but as distributed representations in a neural network, categoryhood must be a gradient concept, so that categories can be created in a gradual manner. Connecting these ideas to the representations of phonology would be an interesting enterprise for the future. A connectionist model that does address phonological issues (Soderstrom, Mathis & Smolensky 2006) unfortunately works with innate constraints (specified in the genome), and therefore, a fortiori, with innate categories (because the constraints refer to phonological categories such as codas); this model therefore cannot handle category emergence.

A separate strand of research involves the modelling of the *perceptual magnet effect* (Kuhl 1991), which is the phenomenon that listeners discriminate two sounds more easily if they belong to different phonological categories than if they belong to the same phonological category; it is as if the auditory properties of two sounds within the same category are nearer to each other than one would expect on the basis of their acoustic distance. Guenther & Gjaja (1996) show with computer simulations that such *perceptual warping* can emerge as the result of the formation of an auditory map in a neural network model. The inputs to the network are auditory values encoded directly as neural activities; for instance, there are one pair of neurons whose activities reflect the second formant (for the first neuron, low activity means low F2, high activity means high F2; for the second neuron, low activity means high F2, high activity means low F2), one pair of neurons whose activity encodes F3, and so on. The model also has a “neural map” consisting of, say, 500 neurons, all of which inhibit each other and all of which are connected to each of the four input neurons. The model is then fed auditory events (F2–F3 pairs) drawn from language-specific distributions; thus, an English language environment is simulated as a Gaussian distribution centred around an F2 of 1000 mel and an F3 of 2075 mel, reflecting the phoneme /l/, plus a Gaussian distribution centred around an F2 of 1000 mel and an F3 of 1200 mel, reflecting the phoneme /ɹ/. As the auditory events come in, a standard learning rule that tries to increase the correlation between presynaptic activity and connection weight for every active cell (a continuous variant of Hebbian learning) causes most cells in the map to become “tuned” to the most frequent combinations of formant values. After learning, a combination of input formants F2–F3 will then generally lead to a different *perceived* combination of formants F2'–F3', if the latter is defined as an average over the “best” tuning frequencies of all active neurons in the map (weighted by their activities); the learning rule will have made sure that the perceived F2'–F3' tends to be close to a frequent combination of input formants, even if F2–F3 are not. This is illustrated in Figure 2.

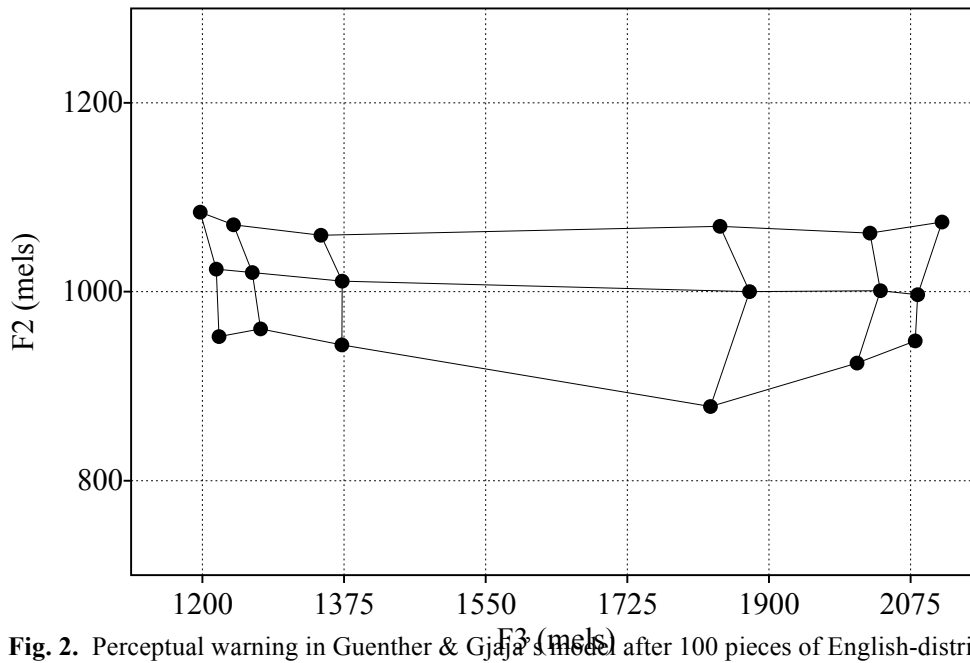


Fig. 2. Perceptual warping in Guenther & Gjaja’s model after 100 pieces of English-distributed input data for liquids. The input formant combinations are the 18 crossings of the dotted lines. The “perceived” formant combinations are the 18 dots.

Whereas Guenther & Gjaja used unrealistically low values for the standard deviations of F2 and F3 (40 and 60 mels, respectively), so that there was essentially no overlap between the formant clouds for /l/ and /ɹ/, Figure 2 was produced with realistic standard deviations (100 and 200 mels, respectively), which required raising the size of a map cell’s “neighbourhood” from 35 to 150 cells (Wanrooij 2009). We can see that for equidistant input formant combinations (the crossings of the dotted lines) the perceived formant combinations (the dots) are no longer equidistant but instead cluster around the centres of the English distributions (F3 = 1200 and 2075 mels; F2 = 1000 mels). If the distance between two dots in the figure is a measure of how well the two sounds can be discriminated, the perceptual magnet effect is explained; for instance, the perceived distance between an input F3 of 1375 and an input F3 of 1550 Hz is reduced to approximately 100 Hz (the distance between the second and third columns of dots in Figure 2), which presumably makes for poor discrimination (“acquired similarity” in terms of Liberman 1957), whereas the perceived distance between an input F3 of 1550 Hz and an input F3 of 1725 Hz is raised to approximately 500 Hz (the distance between the third and fourth columns of dots), which presumably makes for good discrimination (“acquired distinctiveness”, in Liberman’s terms).

A similar result was obtained by Boersma, Escudero & Hayes (2003) with computer simulations of an Optimality-Theoretic learning algorithm: perceptual warping emerged through the use of constraints in favour of perceiving all input F2 and F3 values, constraints *against* perceived F2’ and F3’ values, and constraints *against* perceptual warping. Although both Guenther & Gjaja’s and Boersma et al.’s

simulations are meant to be a part of a larger linguistic model, they would have to rely on a discrete event (a “category creation day”) to turn the warped perceptions into discrete symbolic categories suitable for inclusion in a linguistic model. These models would become more principled if combined with gradual category creation, such as is promised by the distributed connectionist models discussed above.

Finally, there is the promise of *exemplar theory* (Nosofsky 1988), which has been applied to phonological storage by Pierrehumbert (2001) and Wedel (2004, 2006, 2007). This family of theories holds that the lexicon consists of a massive number of stored phonetic (or auditory) events, with or without category labels. Those subtheories that touch on category creation can do so because they include no category labels, but those subtheories that make interesting linguistic generalizations (e.g. on auditory dispersion: Wedel 2006) do require the presence of category labels. Thus, although exemplar theory has the potential of becoming a big theory of language at some point, it cannot yet combine category creation with linguistic theorizing. For instance, exemplar theory cannot handle yet the simplest examples of sentence phonology, such as nasal place assimilation, because it cannot distinguish between underlying forms in the lexicon and surface forms; one could make a version of exemplar theory that includes both surface and underlying forms (Wedel 2004: ch. 4), but even that version cannot handle sentence phonology, because it is incapable of singling out unambiguous underlying forms. It seems that in order to begin to account for basic phonological phenomena such as nasal place assimilation, exemplar theory would have to be extended with stored relations between morphemes and underlying forms, and with relations between underlying and surface forms, thus becoming very much like the model of Figure 1.

A problem that all the above models share (apart from having trouble to link to phonology) is that they rely on the existence of neural mechanisms that can do computations with auditory distance: Guenther & Gjaja’s weighted summation over formant values, Boersma et al.’s distance-dependent anti-warping constraints, and exemplar theory’s nearest exemplars in perception and neighbourhood averaging in production; the underlying networks that should provide such mechanisms are not specified. By contrast, models of associative memory (Kohonen 1984) can derive auditory-distance effects without having auditory distance represented anywhere underlyingly; likewise, there exist Optimality-Theoretic models in which auditory-distance effects emerge without represented auditory distance (for auditory dispersion: Boersma & Hamann 2008), but they do not handle category creation. There still seems to be a divide between models of category creation and models of linguistic processing in many respects.

10.2.8. Conclusion

The conclusion must be that there are no models yet that combine category creation to other emergent properties of language processing, but that some partial answers have been given, so that we may well find a comprehensive model in the future. Such a model may include the linked representations of Figure 1 (plus an articulatory form, as the speaker’s output representation), and represent categories gradiently as distributed across a neurobiologically inspired network, preferably without representing auditory distance explicitly.

References

- Akamatsu, Tsutomu (1997). *Japanese phonetics: theory and practice*. LINCOM Europa, München.
- Berent, Iris, Tracy Lennertz, Paul Smolensky, & Vered Vaknin-Nusbaum (2009). Listeners' knowledge of phonological universals: evidence from nasal clusters. *Phonology* **26**: 75–108.
- Best, Catherine T. (1995). A direct realist view of cross-language speech perception. In Winifred Strange (ed.), *Speech perception and linguistic experience: theoretical and methodological issues*. Baltimore: York Press. 171–203.
- Boersma, Paul (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. PhD thesis, University of Amsterdam.
- Boersma, Paul, Paola Escudero & Rachel Hayes (2003). Learning abstract phonological from auditory phonetic categories: an integrated model for the acquisition of language-specific sound categories. *Proceedings of the 15th International Congress of Phonetic Sciences*. 1013–1016.
- Boersma, Paul (2007). Some listener-oriented accounts of *h*-aspire in French. *Lingua* **117**: 1989–2054.
- Boersma, Paul (2008). Emergent ranking of faithfulness explains markedness and licensing by cue. *Rutgers Optimality Archive* **954**, <http://roa.rutgers.edu>.
- Boersma, Paul (2009). Cue constraints and their interactions in phonological perception and production. In Paul Boersma & Silke Hamann (eds.), *Phonology in perception*. Berlin: Mouton De Gruyter. 55–110.
- Boersma, Paul (to appear). A programme for bidirectional phonology and phonetics and their acquisition and evolution. To appear in Anton Benz & Jason Mattausch (eds.), *Bidirectional Optimality Theory*.
- Boersma, Paul, & Silke Hamann (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* **25**: 217–270.
- Boersma, Paul, & Silke Hamann (2009). Loanword adaptation as first-language phonological perception. In Andrea Calabrese & W. Leo Wetzels (eds.), *Loanword phonology*. Amsterdam: John Benjamins. 11–58.
- Braine, Martin D.S. (1992). What sort of innate structure is needed to “bootstrap” into syntax? *Cognition* **45**: 77–100.
- Carpenter, Gail, & Stephen Grossberg (1987). ART 2: self-organization of stable category recognition codes for analog input patterns. *Applied Optics* **26**: 4919–4930.
- Chomsky, Noam & Morris Halle (1968). *The sound pattern of English*. New York: Harper and Row.
- Clements, G. Nick (1985). The geometry of phonological features. *Phonology Yearbook* **2**: 225–252.
- Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler (1999). Epenthetic vowels in Japanese: a perceptual illusion. *Journal of Experimental Psychology: Human Perception and Performance* **25**: 1568–1578.
- Eisner, Frank (2006). *Lexically-guided perceptual learning in speech processing*. PhD thesis, Nijmegen University.
- Eisner, Frank, & McQueen, James M. (2006). Perceptual learning in speech: stability over time. *Journal of the Acoustical Society of America* **119**: 1950–1953.
- Folkins, John W., & Gerald N. Zimmermann (1981). Jaw-muscle activity during speech with the mandible fixed. *Journal of the Acoustical Society of America* **69**: 1441–1444.
- Fowler, Carol A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* **14**: 3–28.
- Ganong, William F. III (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* **6**: 110–125.
- Grimshaw, Jane (1981). Form, function, and the language-acquisition device. In C.L. Baker & John J. McCarthy (eds.) *The logical problem of language acquisition*. Cambridge, MA: MIT Press. 165–182.
- Grossberg, Stephen (1976). Adaptive pattern classification and universal recoding, I: parallel development and coding of neural feature detectors. *Biological Cybernetics* **23**: 121–134.
- Grossberg, Stephen (1980). How does a brain build a cognitive code? *Psychological Review* **87**: 1–51.
- Grossberg, Stephen (1987). Competitive learning: from interactive activation to adaptive resonance. *Cognitive Science* **11**: 23–63.
- Guenther, Frank H., & Marin N. Gjaja (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America* **100**: 1111–1121.
- Hamann, Silke, Diana Apoussidou & Paul Boersma (to appear). Modelling the formation of phonotactic restrictions across the mental lexicon. To appear in *Proceedings of the 45th Meeting of The Chicago Linguistic Society*.
- Hamann, Silke, Paul Boersma and Małgorzata Cavar (2010). Language-specific differences in the weighting of perceptual cues for labiodentals. *Proceedings of New Sounds 2010*, Poznań. 167–172.

- Harris, John, and Geoff Lindsey (1995). The elements of phonological representation. In Jacques Durand & Francis Katamba (eds.), *Frontiers of phonology: atoms, structures, derivations*. Harlow, Essex: Longman. 34–79.
- Jacquemot, Charlotte, Christophe Pallier, Denis LeBihan, Stanislas Dehaene & Emmanuel Dupoux (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *Journal of Neuroscience* **23**: 9541–9546.
- Jakobson, Roman, Gunnar Fant, & Morris Halle (1963). *Preliminaries to speech analysis*. MIT Press, Cambridge.
- Johnson, Keith, Edward Flemming, & Richard Wright (1993). The hyperspace effect: phonetic targets are hyperarticulated. *Language* **69**: 505–528.
- Jusczyk, Peter (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Keyser, Samuel Jay & Kenneth N. Stevens (1994). Feature geometry and the vocal tract. *Phonology* **11**: 207–236.
- Kohonen, Teuvo (1984). *Self-organization and associative memory*. Berlin: Springer.
- Kuhl, Patricia K. (1991). Human adults and human infants show a “perceptual magnetic effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* **50**: 93–107.
- Liberman, Alvin M. (1957). Some results of research on speech perception: a critical review. *Psychological Review* **72**: 275–309.
- Liberman, Alvin, & Ignatius Mattingly (1985). The motor theory of speech perception revised. *Cognition* **21**: 1–36.
- Liljencrants, Johan & Björn Lindblom (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* **48**: 839–862.
- Lindblom, Björn, James Lubker & Thomas Gay (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics* **7**: 147–161.
- Maye, Jessica & LouAnn Gerken (2000). Learning phoneme categories without minimal pairs. In S.C. Howell, S.A. Fish and T. Keith-Lucas (eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. 522–533.
- Maye, Jessica, Janet F. Werker & LouAnn Gerken (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* **82**: B101–B111.
- McClelland, James L., & Jeffrey L. Elman (1986). The TRACE model of speech perception. *Cognitive Psychology* **18**: 1–86.
- McClelland, James L., & David E. Rumelhart (1986). A distributed model of human learning and memory. In J.L. McClelland, D.E. Rumelhart & the PDP Research Group, *Parallel distributed processing: explorations in the microstructure of cognition. Volume 2: psychological and biological models*. Cambridge, MA: MIT Press. 170–215.
- McGurk, Harry, & John MacDonald (1976). Hearing lips and seeing voices. *Nature* **264**: 746–748.
- McQueen, James M., & Anne Cutler (1997). Cognitive processes in speech perception. In William J. Hardcastle & John Laver (eds.), *The handbook of phonetic sciences*. Oxford: Blackwell. 566–585.
- Norris, Dennis, James M. McQueen, & Anne Cutler (2000). Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences* **23**: 299–370.
- Nosofsky, Robert M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **14**: 700–708.
- Pierrehumbert, Janet (2001). Exemplar dynamics: word frequency, lenition, and contrast. In J. Bybee & P. Hopper (eds.), *Frequency effects and the emergence of linguistic structure*. Amsterdam: John Benjamins. 137–157.
- Pinker, Steven (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pinker, Steven (1989). *Learnability and cognition: the acquisition of argument structure*. Cambridge, MA: MIT Press.
- Polivanov, Evgenij Dmitrievič (1931). La perception des sons d’une langue étrangère. *Travaux du Cercle Linguistique de Prague* **4**: 79–96. [English translation: The subjective nature of the perceptions of language sounds. In E.D. Polivanov (1974): *Selected works: articles on general linguistics*. The Hague: Mouton. 223–237]
- Polka, Linda & Janet F. Werker (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* **20**: 421–435.
- Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Technical Report TR-2, Rutgers University Center for Cognitive Science. [published in 2004 by Blackwell, Malden MA & Oxford]

- Ramus, Franck, Sharon Peperkamp, Anne Christophe, Charlotte Jacquemot, Sid Kouider, & Emmanuel Dupoux (to appear). A psycholinguistic perspective on the acquisition of phonology. In Cécile Fougeron, Barbara Kühnert, Mariapaola d'Imperio & Nathalie Vallée (eds.), *Laboratory Phonology 10*. Berlin: Mouton De Gruyter.
- Samuel, Arthur G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General* **125**: 28–51.
- Saussure, Ferdinand de (1916). *Cours de linguistique générale*. Edited by Charles Bally & Albert Sechehaye in collaboration with Albert Riedlinger. Paris: Payot & C^{ie}. [2nd edition, 1922]
- Savela, Janne (2009). *Role of selected spectral attributes in the perception of synthetic vowels*. Ph.D. dissertation, University of Turku.
- Slobin, Dan I. (1997). The origins of grammaticizable notions: beyond the individual mind. In Dan I. Slobin (ed.), *The crosslinguistic study of language acquisition: expanding the contexts*. Mahwah: Lawrence Erlbaum.
- Smolensky, Paul (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry* **27**: 720–731.
- Soderstrom, Melanie, Donald Mathis, & Paul Smolensky (2006). Abstract genomic encoding of Universal Grammar in Optimality Theory. In Paul Smolensky & Géraldine Legendre, *The harmonic mind*. Cambridge, MA: MIT Press. 403–471.
- Steriade, Donca (2001). Directional asymmetries in place assimilation. In Elizabeth Hume & Keith Johnson (eds.), *The role of speech perception in phonology*. San Diego: Academic Press. 219–250.
- Tomasello, Michael (2003). *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA & London: Harvard University Press.
- Wanrooij, Karin (2009). Does distributional input improve the categorization of speech sounds? Neurobiological aspects and computer simulations. MA Thesis, University of Amsterdam.
- Wedel, Andrew (2004). *Self-organization and categorical behavior in phonology*. PhD thesis, University of California at Santa Cruz.
- Wedel, Andrew (2006). Exemplar models, evolution and language change. *The Linguistic Review* **23**: 247–274.
- Wedel, Andrew (2007). Feedback and regularity in the lexicon. *Phonology* **24**: 147–185.