

The emergent paradigm in Laboratory Phonology: Phonological categories and statistical generalisation in Cutler, Beckman and Edwards, Frisch and Bréa-Spahn, Kapatsinski, and Walter

KAREN CROOT

University of Sydney

Abstract

This commentary responds to Abigail Cohn's review of successes and challenges in Laboratory Phonology at the LabPhon10 meeting in 2006 (Cohn 2010), and her call for "truly integrated" theoretical models. In my view, the sort of integrated theory that Cohn endorsed relies on the availability of a scientific paradigm to organise our approach to data, methodology and theory. I propose that, from its codification as an approach by the first LabPhon meeting in 1987 (Beckman and Kingston 1990), research in Laboratory Phonology has been (and still is) in the process of acquiring such a paradigm. My commentary argues that the occurrence of categorisation at various levels of granularity, together with evidence for gradient category membership at all levels, are central findings in Laboratory Phonology. These sit in the paradigm alongside the core research question about how the physical, cognitive and social properties of speech sounds are related, and a commitment to using hybrid methodologies from phonetics and phonology, computational linguistics, psycholinguistics and sociolinguistics. My commentary highlights these aspects of the emerging LabPhon paradigm in the papers in this issue by Cutler, Beckman and Edwards, Frisch and Bréa-Spahn, Kapatsinski, and Walter.

Cohn (2010) defined Laboratory Phonology as an approach to investigating human sound systems using integrated methodologies within various (or no) theoretical frameworks: a scientific enterprise in need of integrated models to account for rich and varied empirical data. My commentary will discuss the notion of a scientific paradigm as proposed by Kuhn (1970), and conclude that from the outset, research in Laboratory Phonology has been in the process of moving from pre-paradigm "fact-gathering" to the practice of "normal science" guided by a paradigm. The emphases on understanding the relationships between the phonetics and phonology, the physical and cognitive aspects of speech sounds (Beckman and Kingston 1990, Cohn 2010), and on the use of hybrid methodologies, have been central to research in Laboratory Phonology from the start. Factors influencing the development of

the paradigm include the LabPhon community's shared interest in mind-matter relationships, the development of theories integrating hierarchical data structures on different time scales, and the investigation of human sound systems in the context of larger systems. Two research findings that are proving central to the emergent paradigm are the occurrence of *categorisation at various levels of granularity*, and the evidence for *gradient category membership at all levels*.

1. Integrated models require a scientific paradigm

In the terms of analysis offered by historian of science Thomas Kuhn (1970), Cohn's (2010) paper expressed the need for research in Laboratory Phonology to occur within a *paradigm*: a conceptual framework that identifies and integrates the most important issues within a scientific field. Kuhn distinguished the practice of science as being qualitatively different in a *pre-paradigm* compared with a *post-paradigm* period. Pre-paradigm research is characterised by "fact-gathering" by different "schools" within a discipline, in which "different men confronting the same range of phenomena, but not usually all the same particular phenomena, describe and interpret them in different ways" (Kuhn 1970: 17). Kuhn (1970: 13–14) gives the following example of pre-paradigm investigation of electricity:

"During that period there were almost as many views about the nature of electricity as there were important electrical experimenters. . . . All their numerous concepts of electricity had something in common – they were partially derived from one or another version of the mechanico-corpuseular philosophy that guided all scientific research of the day. In addition, all were components of real scientific theories, of theories that had been drawn in part from experiment and observation and that partially determined the choice and interpretation of additional problems undertaken in research. Yet though all the experiments were electrical, and most of the experimenters read each other's works, their theories had no more than a family resemblance."

I cast my eye over the programme for the LabPhon11 meeting and asked myself to what extent our current investigation of phonological representation could appropriately be described as fact gathering by different schools. I wondered whether, with some judicious editing, the description of pre-paradigm research on electricity might serve as a helpful parable about progress in our own field:

" . . . there were almost as many views about the nature of *phonological representations* as there were important *Laboratory Phonologists*. . . . All their numerous concepts of *phonology* had something in common – they were partially derived from one or another version of the *generative and information-processing* philosophies that guided all *linguistic* research of the day. In addition, all were components of real scientific theories, of theories that had been drawn in part from experiment and observation and that partially determined the choice and interpretation of additional problems undertaken in research. Yet though all the experiments were *about phonological representations*, and most of the experimenters read

each other's works, their theories had no more than a family resemblance" [words in italics my modifications of Kuhn 1970: 13–14].

According to Kuhn (1970), pre-paradigm fact-gathering may appear random or yield a "morass" of information that is hard to integrate. This contrasts with post-paradigm "normal science" characterised by a common body of belief taken for granted, shared assumptions and values, agreement about important previous findings, and a group-licensed ability to see a variety of situations as similar. A *paradigm* guides the whole group's research by permitting the selection, evaluation and criticism of data, theory, and methodology, and is more successful than alternative views in solving certain problems that scientists in the field have come to recognise as acute. Within a paradigm, central findings and the primary thrust of research are frequently driven by knowledge that can be compressed into a few key formulae or general "laws". A paradigm does not, however, require scientists to hold shared theories or models. These are too specific. Instead, there can be many theories and models, each framing the high-level assumptions of the paradigm in a manner that allows them to be tested in a specific context.

Pierrehumbert and Clopper (2010: 116–119) wrote of three "shared commodities in a free-trade zone" possessed by the LabPhon community: (i) a shared interest in mind-matter relationships (the "relationship between the physical reality of speech and the cognitive representation of language"), (ii) autosegmental-metrical phonological theory, which incorporates "hierarchical data structures from different time scales . . . analysed in a single theoretical framework", and (iii) a recent interest in the "relationship between levels of representation and the social function and social context of language". Below I show how these "commodities" in fact reflect key aspects of the emerging LabPhon paradigm.

2. Mind-matter relationships

The Labphon community's interest in mind-matter relationships highlights the importance of embodiment, of speaker-listener behaviour firmly set in the context of the physical body – the sensory organs and motor system – and the context of the task, and the physical and linguistic environment (Steels and de Boer 2007). Laboratory Phonology has previously taken extensive account of progress in the disciplines of experimental acoustic and articulatory phonetics and psycholinguistics. However, perhaps more important is LabPhon's recent consideration of human learning systems, because of the potential this offers for integrating data across these broad sources. According to Kuhn (1970), a paradigm emerges when initial divergences begin to disappear due to the triumph of one of the pre-paradigm schools, simply, and often serendipitously, because they happen to emphasise one part of the data that turn out to be central in explaining all the others. Two related characteristics of the data that we increasingly recognize as central in Laboratory

Phonology are the sorts of linguistic *categories* so effectively identified and analysed using verbal/symbolic theories, and the *gradient* that appears at all levels of analysis: the probabilistic nature of sound structures (Pierrehumbert 2001). The emerging recognition of the importance of these characteristics coincides with the development of computational technologies that allow us to simulate and conceptualise how such structures might be learned.

There are now connectionist and statistical learning models in many areas of interest to Laboratory Phonology: the Minimal Generalisation Learner (MGL, Albright and Hayes 2003), Shortlist (Norris 1994), TRACE (McClelland and Elman 1986), WEAVER (Roelofs 1997), DIVA (Guenther 2002), the Dell models (e.g. Dell et al. 1997) and a range of others. In his commentary at LabPhon5, Dell (2000: 347) explained how connectionism, which accounts for the majority of these models, yields a number of what he called “nice effects for free”. Important properties of connectionism are spreading activation and learning, representations distributed over many units and connections, and outputs that are sensitive to the structure of the input and constraints on the task. Gradient, overlapping categories emerge, as a function of learning, as patterns of shared activation over the processing units of the model (for example, categories such as “bird” in the model of semantic knowledge of Rumelhart and Todd 1990, described by McClelland and Rogers 2003). Dell (2000: 348) concluded, “the kinds of theories that benefit most from a connectionist perspective are those that emphasise the role of learning and recent experience, graded rules, constraint satisfaction, and how knowledge is used in actual tasks.” Perhaps the “nicest effect” for Laboratory Phonology is that the properties that distinguish statistical learning approaches from verbal/symbolic approaches correspond to some of the acute problems that a paradigm for Laboratory Phonology must address.

2.1 *Abstract phonological categories . . . but not as we know them*

Given LabPhon’s emphasis on the importance of gradient category membership and the development of connectionist models that begin to show us how the human language system might extract structure from linguistic data, I turn to the first question I was asked to address in this commentary: Is there evidence for or against the existence of abstract phonological categories? There is evidence for multiple levels of structure, including but not restricted to traditional phonological categories.

Walter reports on the observed versus expected co-occurrence of vowels compared with consonants in Croatian and Spanish, finding stronger cooccurrence restrictions for consonants than vowels, with similar vowels actually preferred in Croatian. This is evidence for a category of vowels that behaves differently to a category of consonants with respect to co-occurrence avoidance in these two languages, and in other languages reviewed by Walter. Walter notes, however, that even such apparently fundamental categories as vowels and consonants are somewhat permeable.

Cutler describes abstract representations approximately corresponding to phonemes, that allow generalisation of learning about unusual fricative production to untrained words containing those fricatives. The specificity of the learning in these experiments compels a learning mechanism that is sensitive to speaker characteristics as one aspect of the environmental structure extracted. The fact that there is also lexical generalisation demonstrates that information about structure must be stored, rather than computed online. However, Cutler also refers to work by Kraljic and Samuel (2006) that suggests that fricative category information is speaker-specific whereas stop category information can generalise to new speakers. This is evidence for categories similar to traditional manner class categories, with fricatives and stops behaving differently, but these fricative and stop categories are functionally dissimilar in a particular way that current phonological theories do not capture.

Beckman and Edwards discuss some of the difficulties they found in developing equivalence classes that captured articulatory similarity across languages, and noted the need for finer-grained distinctions than phonemic categories to capture the features that make consonants equivalent. Thus the structures needed to analyse multilingual data about the developing lexicon also do not perfectly correspond to traditional phonological categories.

2.2 *A climbing wall of abstraction*

A second question I was asked to address was: What types of statistical generalisations emerge specifically from the lexicon? Pierrehumbert (2003: 178) suggested that viable theories of human sound systems will contain “a ladder of abstraction, with each level having its own representational apparatus”. I suggest that, instead, our theories may need to contain something more like a climbing wall than a ladder. In this conceptualisation, chunks and lumps (probabilistic categories) at different levels of granularity, and with different degrees of overlap, would be distributed over the representational space, emerging partly as generalisations over the lexicon(s) of the language(s) a person acquires.

A number of statistical generalisations were shown to emerge from the lexicon in the papers in this session. Frisch and Bréa-Spahn report that well-formedness judgments for nonwords reflected self-rated size of vocabulary for English words. They propose that phonotactic knowledge emerges from generalisations over an individual’s unique linguistic experience. Cutler argues that generalisations about prosodic structures are also learned. Using simulations and human artificial language learning, Kapatsinski reports that competition between possible input:output mappings is resolved by comparison of the relative reliability of mappings, defined as the number of words to which the rule applies divided by the number of words to which it could apply. He shows that velar palatalisation rules estimated over the whole Russian lexicon do not know about the structure of English loan words, so although highly reliable for Russian, they fail for velars in English loan words.

Another type of structure extracted is morphological alternation correlated with differing segmental content in stems. A strength of the MGL (Albright and Hayes 2003) simulations is that they allow competition between generalisations at different levels.¹ The inability of the MGL to account for non-local effects (e.g. vowel harmony) shows that the eventual learning mechanism will be based on correlations that are more complex than those obtained from linear strings of features, segments and so on.

Beckman and Edwards found some, but not all, of the correlations they predicted between articulatory complexity, frequency within and across languages, and production accuracy for consonants, in 2–3 year olds acquiring Cantonese, English, Greek and Japanese. They conclude that consonants that would be classified as “the same” under traditional phonological theory are in fact based on different cognitive representations in children acquiring different languages because such representations emerge from individual children’s growing lexicons. Thus a major source of input to a child’s developing language system is the lexicon of the language he or she is acquiring, and the child’s statistical learning mechanisms extract structure present in that lexicon. However, effects of structure in the adult lexicon of the language upon children’s developing phonological representations will be qualified by differences in the lexicons of child- versus adult-directed speech, and by differences in the developing lexicons of one child versus another.

3. Mapping between qualitatively different data structures

3.1 Hierarchical data structures on different time scales

The second commodity identified by Pierrehumbert and Clopper (2010) was the use of autosegmental metrical phonological theories in which phonological representations are multidimensional, consisting of independent tiers linked to each other. The most important feature of such theories is that they are able to represent hierarchical data structures on different time scales. Further, we can compare an *autosegmental metrical model* linking segmental and metrical data structures to a *connectionist model* mapping between, for example, acoustic-phonetic units, articulatory units and lexical-semantic units (e.g. Plaut & Kello 1999). The crucial factor in both types of model is that mapping between different types of information allows the integration of qualitatively different data structures. This question of how to map between qualitatively different data structures (phonetic/phonological, physical/cognitive etc.) has been central to research in Laboratory Phonology from the start, and is evident in all the papers I review here.

Walter’s paper raises a question about how qualitatively different types of information might be integrated. She discusses several reasons why vowels might behave differently to consonants in regard to expected co-occurrence patterns. These include the fact that consonant place of articulation conveys more lexical informa-

tion than vowel height and fronting, differences in articulatory effort, the increased computational resources required to track vowel compared with consonant transition probabilities, and the fact that vowels convey more information on the F_0 channel, including speaker-indexical information. To tease out the separate contributions of these factors, one would need to consider the influence of ambient language, articulatory properties, cognitive resources, and indexical information – so one would need the type of model that can integrate these different structures.

In order to map between qualitatively different data structures, we also need to know the limits of each factor. The papers of Cutler, Walter, and Beckman and Edwards all report cross-linguistic work that can tell us to what degree the probabilistic structures used in language processing arise due to the structure of a particular language, and thus, how much has still to be explained. Frisch and Bréa-Spahn go a step further by investigating the relative contributions of two languages in the same (i.e. bilingual) speakers. Kapatsinski applies the MGL previously used to model English (Albright and Hayes 2003) to investigate properties of Russian.

Mapping between different data structures may also entail consideration of the timing of operations within and between structures. The papers by Cutler and Kapatsinski draw conclusions about the sequencing of operations within a particular area of analysis. Cutler notes that lexical distributions provide abstract probabilities that guide segmentation and that provide candidates for lexical recognition after segmentation. Kapatsinski argues for a critical effect of penultimate segment only when the stem extension is chosen at the same time as the decision to palatalise the suffix. Outside of papers in this session, there is work on entrainment between systems (e.g. Barlow and Estep 2006) and on the relative timing of processing of the sorts of information represented on independent tiers of an autosegmental model (e.g. prosodic versus segmental information, Keating and Shattuck-Hufnagel 2002). When the LabPhon community begins to use information about brain function to constrain theories of sound structure, time-course information can be obtained from event-related potential (ERP) and magnetoencephalography (MEG) methodologies.

3.2 *Evolution, history, culture, cognition – and Laboratory Phonology*

The final commodity described by Pierrehumbert and Clopper (2010) was the need to consider relationships between levels of representation and the social functions and contexts of language. This is a new-ish interest within Laboratory Phonology, and only one of the ways in which speech and the representation of language are being conceptualised as participating in bigger systems and being constrained by them. The factors influencing phonological representation are located both within and outside the individual, further underlining the LabPhon community's core interest in mapping between qualitatively different data structures as described above.

In terms of conceptualising speech and language within larger systems, Cutler describes the categories used in language perception tasks (e.g. phonetic

categorisation, segmentation of continuous speech using prosodic probabilities) in the context of human categorisation more generally (e.g. colour categorisation, visual letter recognition). She concludes that we should understand speech perception as one application of a very general property of human learning systems. Beckman and Edwards locate their investigation of children's developmental consonant misarticulations in the context of biological evolution and the evolution of languages and cultures: broad explanatory systems indeed.

4. The road to a firm research consensus

Do the “nice effects” such as multiple levels of structure and gradient representations that emerge from connectionist architectures mean we need to do away with verbal/symbolic theories in Laboratory Phonology? Certainly not – we continue to work with them for several reasons. First, the categories within traditional phonological theories have high degrees of descriptive and predictive power, and second, we cannot all run connectionist simulations. Most importantly, such theories are essential to refine assumptions for statistical learning models – and these models are only as good as their assumptions about input/output representations and architecture (Dell 2000). At the same time, however, we need to acknowledge that verbal/symbolic theories involving discrete categories have limited ability to capture gradience in data and to make quantitative predictions. So there is a place for both statistical learning models and verbal/symbolic theories, and they can be complementary.

A second question is whether theories based on statistical learning mean we have to do away with innate constraints in our theories of phonological representation and processing? Again, the answer is no, but the sorts of constraints we end up positing may look different. According to Chater and Manning (2006), probabilistic learning does not necessarily entail a totally empiricist system. As well as being able to model what people are able to learn, statistical learning models can also tell us when constraints *must* be innate because they *cannot* be learned, or at least do not appear to emerge from environmental learning.

In summary, to understand the relationship between the cognitive and physical aspects of human sound systems, we need to understand relationships between human auditory, articulatory and learning systems, linguistic and indexical/speaker-specific data in the speech signal, performance factors (attention, executive processing), task demands, and social context – and it goes without saying that such relationships are exceptionally complex. However, our pre-paradigm fact-gathering and our shared attention to solving the problems inherent in studying these relationships per sé have brought Laboratory Phonology close to a paradigm in which categorisation and gradient category membership are recognized as central phenomena to be explained. As in other areas of cognitive science, we can continue to investigate the phenomena of interest using verbal/symbolic theories and statistical learning approaches.

According to Kuhn (1970: 15), “History suggests that the road to a firm research consensus is extraordinarily arduous.” Instead, with continued emphasis on the value of LabPhon as a nurturing community of scientists, and on science as a social endeavour,² we might even enjoy the journey.

Correspondence e-mail address: karenc@psych.usyd.edu.au

Notes

1. Thank you to Stefan Frisch for pointing this out.
2. Thank you to Louis Goldstein for his comment to this effect following my spoken paper.
3. Thank you to Paul Warren and Jen Hay for their helpful comments on an earlier version of this manuscript.

References

- Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90. 119–161.
- Barlow, Steven M. & Meredith Estep. 2006. Central pattern generation and the motor infrastructure for suck, respiration, and speech. *Journal of Communication Disorders* 39(5). 366–380.
- Beckman, Mary E. & John Kingston. 1990. Introduction. In John Kingston & Mary E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, 1–16. Cambridge: Cambridge University Press.
- Chater, Nick & Christopher D. Manning. 2006. Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences* 10(7). 287–291.
- Cohn, Abigail C. 2010. Laboratory Phonology: Past successes and current questions, challenges and goals. In Cécile Fougeron, Mariapaola D’Imperio, Barbara Kühnert & Nathalie Vallée (eds.), *Variation, Detail and Representation: Laboratory Phonology X*. Berlin: Mouton de Gruyter.
- Dell, Gary S. 2000. Commentary: Counting, connectionism, and lexical representation. In Michael B. Broe & Janet B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 335–348. Cambridge: Cambridge University Press.
- Dell, Gary S., Myrna F. Schwartz, Nadine Martin, Eleanor M. Saffran & Deborah A. Gagnon. 1997. Lexical access in aphasic and nonaphasic speakers. *Psychological Review* 104. 801–838.
- Guenther, Frank H. 2003. Neural control of speech movements. In Antje Meyer & Neils Schiller (eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, 209–239. Berlin: Mouton de Gruyter.
- Keating, Patricia & Stephanie Shattuck-Hufnagel. 2002. A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics* 101. 112–156.
- Kraljic, Tanya & Arthur G. Samuel. 2006. Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review* 13(2). 262–268.
- Kuhn, Thomas. 1970. *The structure of scientific revolutions* (Second Ed). Chicago: The University of Chicago Press.
- McClelland, Jay L. & Jeff L. Elman. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18. 1–86.
- McClelland, Jay L. & Timothy T. Rogers. 2003. The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience* 4. 310–322.

- Norris, Dennis G. 1994. Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52(3), 189–234.
- Pierrehumbert, Janet B. & Cynthia G. Clopper. 2010. What is Labphon? And where is it going? In Cécile Fougeron, Mariapaola D’Imperio, Barbara Kühnert & Nathalie Vallée (eds.), *Variation, Detail and Representation: Laboratory Phonology X*, 113–130. Berlin: Mouton de Gruyter.
- Pierrehumbert, Janet B. 2001. Stochastic phonology. *GLOT* 5(6), 1–13.
- Pierrehumbert, Janet B. 2003. Probabilistic Phonology: Discrimination and Robustness. In Rens Bod, Jen Hay & Stefanie Jannedy (eds.), *Probability Theory in Linguistics*, 177–228. Cambridge MA: MIT Press.
- Plaut, David C. & Christopher T. Kello. 1999. The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In Brian MacWhinney (ed.), *The emergence of language*, 381–415. Mahwah, NJ: Erlbaum.
- Roelofs, Ardi. 1997. The WEAVER model of word-form encoding in speech production. *Cognition* 64(3), 249–284.
- Rumelhart, David E. & Peter M. Todd. 1993. Learning and connectionist representations. In David E. Meyer & Sylvan Kornblum (eds.), *Attention and performance XIV*, 3–30. Cambridge, MA: MIT Press/Bradford Books.
- Steels, Luc & Bart de Boer. 2007. Embodiment and Self-organization of Human Categories: A Case Study for Speech. In Jordan Zlatev, Tom Ziemke, Roz Frank & René Dirven (eds.), *Body, Language, and Mind* (Volume 1), 241–259. Berlin: Mouton de Gruyter.