J. Linguistics **47** (2011), 201–217. © Cambridge University Press 2011
doi:10.1017/S002222671000040X

REVIEW ARTICLE

A history maker[1]

JOHN COLEMAN

Phonetics Laboratory, University of Oxford

(Received 2 June 2008; revised 1 December 2010)

Maria-Josep Solé, Patrice Speeter Beddor & Manjari Ohala (eds.), *Experimental approaches to phonology* (Oxford Linguistics). Oxford: Oxford University Press, 2007. Pp. xviii + 465.

1. Introduction

The systematic positioning of phonology as an almost entirely separate discipline from phonetics by members of the Prague School in the 1930s (see e.g. Cercle Linguistique de Prague 1931) had several regrettable consequences. Considering phonological units to be psychological, not physical, attention to the detailed substance of speech seemed irrelevant; experimental phonetics was afforded a place, but that place was on the periphery, as an AUXILIARY discipline of linguistics. Largely divorced from the investigative methods of cognitive psychology, many of which arose later than structural linguistics anyway, phonologists have amassed a body of analyses and modes of enquiry rooted in traditional grammatical discourse. The 'arbitrariness of the sign' may, if we are incautious, too easily lapse into a general arbitrariness of linguistic analysis. If we ask why high vowels are lowered to mid in stressed closed syllables in Chamorro (Chung 1983: 37) whereas low and mid vowels are raised when unstressed in Bulgarian (Trubetzkoy 1939 [1969]: 81; Klagstad 1958: 46; Jakobson 1962: 446), a phonological theorist may answer that there is such-and-such a rule, or such-and-such a combination of constraints, and consider this to be an explanation. In spite of the claim of linguistics to be the science of language, quantitative, measurement-based experiments are almost entirely absent from phonological theory and considerations of methodology are often regarded as an irrelevant distraction.

[1] I thank Nigel Fabb and an anonymous *JL* referee for their trenchant – and thus extremely helpful – comments on the manuscript. Of course, I am also grateful for the other referee's approval of the paper! And to Larry Hyman for a couple of very agreeable discussions about experimental vs. notebook methods in phonology, which taught me to better appreciate his point of view and to refine my language a little.

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

The plausibility of a result, or its simplicity, or some other aesthetic property, is considered more important than how it was arrived at, an attitude that may engender a distrust of or at best a disregard for experimental methodology.

Experimental phoneticians and psychologists of language have nevertheless continued to 'do their thing', amassing a distinct body of discoveries concerning phonological patterns in speech and what speakers and listeners know of them. In addition, sociolinguists and functionalist-inclined phonologists have sought extra-grammatical reasons for observed sound patterns, and historical changes to those patterns. Steadily, such work has accumulated to the point that experimental phonology has gradually carved out a place as subculture of linguistics, for which some parts of the volume under review (e.g. the chapters by Keith Johnson and Juliette Blevins) are recommendable as essential reading even for NON-phonologists, as they challenge long-ingrained views: the introductory textbooks in linguistics and phonology will need to be rewritten. But more of that later.

John J. Ohala has done more than anyone else to bridge the disciplinary gulf between the old-school descriptive and theoretical phonologists, on the one hand, and experimentalists, on the other. This volume was written in his honour, and many of its twenty three chapters came from papers given at a conference to fête him held in Berkeley, California, in May 2004. They are arranged in five parts: I. Theory and background, II. Phonological universals, III. Phonetic variation and phonological change, IV. Maintaining, enhancing, and modeling phonological contrasts, and V. Phonotactic and phonological knowledge. Rather than attempting to distil the contents of 23 self-standing chapters into as many brief summaries, I focus in Section 2 on five areas of significant disagreement between the 'experimental' and 'theoretical/symbolic' approaches to phonology. This is not to deny that there are very many aspects of phonology in which experimental results support or are at least consistent with the claims and understanding of non-experimental, theoretical phonology; it's just that there is more to learn by focussing on the disagreements. But to reach a less confrontational conclusion, in Section 4 I shall also consider one emerging area of work in which descriptive phonology and experimental work clearly converge: the development of corpus-based methods. And in recognition of the fact that work by experimentalists is certainly fallible, in Section 3 I shall discuss some chapters whose criticisms of symbolic phonology are rather off-target.

2. EXPERIMENTAL VS. THEORETICAL PHONOLOGY

Five serious disagreements between experimental phonology and theoretical work are repeatedly exemplified in this volume:

1. *The importance of method.* While generative phonologists, in particular, have taken a strong line against the importance of method (cf. the

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

rejection of 'discovery procedures'), experimentalists care deeply about method and consider methodological development to be important.

2. *The place of statistics*. Are the units or processes of phonology inherently stochastic, or does statistics enter as merely an adjunct to experimental method?
3. *Timing*. Time is treated in theoretical phonology in terms of discrete, countable units (segmental slots, moras, etc.), whereas in experimental work, time is time, a continuous quantity measured in e.g. milliseconds.
4. *Acoustic vs. articulatory categories*. There has been an ongoing tension within phonology as to whether distinctive features and natural classes partition along fault-lines that are shaped by or are at least best described in acoustic or articulatory terms. Since Chomsky & Halle (1968), articulatory categories have dominated the discourse; nevertheless, a number of works in the experimental tradition seem to support the retention of certain acoustic categories.
5. *Abstract symbols vs. detailed examples*. Storage of numerous phonetically concrete examples of word forms has for decades been regarded by theoreticians as rather an implausible non-starter. In recent years, however, 'exemplar' models of phonological representation have derived support from some experimental results.

2.1 *The importance of method*

The volume's honorand, John Ohala, provides the opening piece, 'Methods in phonology', arguing that a scientific discipline is characterized by 'the questions it asks; the answers given to the questions, that is, hypotheses and theories; the methods used to marshal evidence in support of the theories' (3). It is in this third aspect that John Ohala and colleagues depart from e.g. Halle's (1959: 12) view that 'the methods by which a scientific description are discovered are not of essential concern to the science in question', i.e. Chomsky's classic rejection of the search for 'discovery procedures'. While it is certainly true that discoveries and the nature of the evidence for them are more important than the means or process by which they are discovered, it is also true that the questions a science asks are in turn shaped by WHAT CAN BE MEASURED, AND HOW WELL, which depends on the techniques of the day. John Ohala notes that developments in method 'constitute the principal engine for refinement and productive change in a discipline, helping to moderate the pace with which one theory supplants another' (4–5); he mentions the refinement of the kymograph by Marey in the nineteenth century, which made it possible to record relatively rapid physiological events (vocal cord vibration, for example).

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

To John Ohala's selection of examples let me add a few more. First, the invention of the sound spectrograph enabled[2] the discovery of 'gravity', the low second formant frequency that is the common acoustic signature of velar and labial consonants, and back and rounded vowels (Jakobson, Fant & Halle 1951). Without this knowledge, the linguistically natural and frequent association of velarity with labiality (as in double articulations such as [w], [u], [g͡b], and a host of sound changes such as [x] > [f]) lacked an explanation. Second, the subsequent invention of 'pattern-playback' formant synthesis from tracings of spectrograms paved the way for the discovery of the categorical perception of place of articulation of plosives (Cooper et al. 1952, Liberman et al. 1967). The diversity of acoustic cues to alveolar place of articulation led Liberman 1957 to propose the Motor Theory of Speech Perception, a spur[3] to the move from acoustically-based distinctive features to articulatory distinctive features made by Chomsky & Halle 1968. Third, the technology for making permanent physical recordings of sound and movement allows such data to be repeatedly reexamined, measured, and accumulated into large databases, such as the Switchboard corpus (Godfrey, Holliman & McDaniel 1992) or the Kiel Corpus of Spontaneous Speech noted in Klaus J. Kohler's chapter.

Phonological experiments may examine speech, but equally they may examine subjects' responses to speech, or behaviour in a speech-related task. In 'What's in CVC-like things? Ways and means to look at phonological units across languages', Bruce L. Derwing reminds us of some of the discoveries of psycholinguistics, and presents some new data from Minnan (Southern Min) Chinese. I cannot agree with his pessimism that 'it seems highly unlikely that any of these higher-order theoretical entities [segments, syllables, morphemes, word, lexical categories etc.] are ever going to be found even in the "wetware" of a physical brain, much less as part of the physical "stream of speech"' (327): progress on the 'wetware' front – albeit still at a very early stage – is exemplified at a variety of linguistic levels by e.g. Caramazza et al. 2000 (double dissociation of errors in producing consonants vs. vowels), Damasio & Tranel 1993 (double dissociation of errors in retrieval of nouns *vs.* verbs), Just et al. 1996 (recruitment of cortical regions in processing embedded clauses), and a steady stream of results using brain imaging methods. In the 'physical stream of speech' work in speech and language technology and its application by some to linguistics shows that units of various linguistic categories can be found if one looks (with a

[2] Or rather, *provided supporting evidence*: the terms 'grave' and 'acute' were used to describe back vs. front vowel sounds by Wolfgang von Kempelen as early as 1791, though he experimentally challenged the hypothesized contrast (see excerpts at http://www.phon.ox.ac.uk/jcoleman/vonKempelen.html). The terms were widely used by Eichhoff 1836, Jakobson 1939 and others prior to the first publications on the sound spectrograph.

[3] Liberman's proposal was taken up by Halle & Stevens 1964.

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

theory, of course), such as distinctive features (Coleman 2003), segments (Stevens 2002), and from transcriptions of speech may be discovered morphemes (Goldsmith 2001), parts of speech (Rapp 2007) and so on into other aspects of syntax and semantics. Nevertheless, there is no disagreement that knowledge of language can be and is being appropriately studied through psycholinguistic experiments; this is the one chapter that I would be most likely to set linguistics undergraduates to read. Derwing and co-workers' studies of syllable structure in the minds of Korean and Minnan speakers, including preliterate schoolchildren, clearly demonstrate the language-specific psychological reality of a body + coda (CV.C) syllable structure in those languages, in contrast with the onset + rime (C.VC) syllable structure of English. As I said above, introductory linguistics textbooks will need to be rewritten.

2.2 *The place of statistics in phonology*

Some textbooks present experimental inference in the scientific method as a fairly straightforward application of logic of the kind: 'If theory A, then B should happen. B didn't happen, therefore theory A is rejected/not supported'. In reality, statistics and probability theory are indispensable features of experimental inference and argumentation in all sciences, whereas the categories and rules of phonological theory are standardly discrete, not probabilistic or gradient. Statistics is essential in experimental phonology because of endemic variation in everything human. Because an effect might have multiple interacting causes, we begin by attempting to isolate and manipulate factors individually, by holding all other variables constant. For example, when using human subjects, we may regard it as important to use a group that is homogenous with regard to age group, sex, social background, language, dialect, religion, level of education, whether or not they are trained in phonetics/linguistics/psychology, time of day (people get more tired as the evening passes), whether they smoke, use drugs, are on medication, etc. The variables studied are quantitative i.e. counts or measurements. A single set of measurements may be faulty (as a result of inadvertent error), or even if accurate, unrepresentative of other situations, and hence not a suitable basis on which to make general statements. Additionally, natural phenomena are often inherently variable, to a certain degree. We would like to understand and account for that variability, rather than wish it away. To avoid the fragility of single observations, experiments are usually repeated many times e.g. through the use of several subjects, re-runs of the procedure, reproducing the work of other experimenters to confirm or refute their conclusions. Because of multiple trials, a lot of data may be collected germane to a single question. Statistical techniques are then necessary to determine the significance of the mass of data to the question under consideration. Advances in methodology, therefore, may include advances in statistics.

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

Slips of the tongue are a source of data regarding phonological structure and language processing that is often mentioned anecdotally but only rarely elicited experimentally. In 'The SLIP technique as a window on the mental preparation of speech: Some methodological considerations', Sieb Nooteboom & Hugo Quené show us in some detail why it is difficult to obtain sufficient data for such experiments to be effective. A particular problem is that because even elicited slips are quite rare, a large number of subjects needs to be recruited; however, since previous experiments have found that subjects may behave differently, a large number of subjects will introduce between-subject variability, with attendant statistical problems. Nooteboom & Quené propose a new statistical approach (new to this experimental paradigm, that is), multi-nomial logistic regression, applying it to some earlier data, avoiding spuriously significant effects and revealing a new account of what causes the lexical bias in phonological speech errors.

Phonological theorists may be comfortable to allow statistics as a necessary tool of experiments, while maintaining that phonological knowledge is symbolic, not variable or statistical. Some recent work, however, advances the proposition that linguistic knowledge may sometimes be statistical (Saffran, Aslin & Newport 1996, Bybee 2001, Bonatti et al. 2005). For example, one might regard markedness as simply a veiled surrogate for usage frequency. Debate on this question is taken up in Section 2.5 below.

2.3 *Timing*

Many of the papers in this volume depend on the accurate measurement of TIME in acoustic or physiological records; measurement vs. disregard of time is one of the hallmark differences between experimental vs. theoretical phonology.

In 'Articulatory movements and phrase boundaries', Patrizia Bonaventura & Osamu Fujimura seek the main factors responsible for the larger and longer articulatory movements associated with prosodic phrase edges (e.g. phrase-final lengthening). In the terms of Fujimura's C/D model of articulatory control, they examine the relative role played by the magnitude and speed of articulator excursions, and of the size of the 'gap' between syllables – a theory-internal concept that is inferred rather than directly observed, albeit from fleshpoint movement data obtained using X-ray microbeam They find that excursion (i.e. the amount of movement) contributes most to the prediction of speed; syllable duration and boundary magnitude have a lesser effect, for some speakers. Fujimura's earlier claim of a within-speaker, invariant 'iceberg' pattern of velocity vs. time for articulator movements requires amendment in the light of this new data: iceberg patterns are invariant after removing a predictable linear effect of excursion.

In 'Controlled and mechanical properties in speech: A review of the literature', Maria-Josep Solé surveys a body of work, to which she has made

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

key contributions, in which variation in the fine temporal structure of speech due to changes in speech rate or stress is seen to vary between languages, providing tell-tale information about which aspects of speech can or cannot be under the control of the speaker's nervous system and conversely, which aspects of timing are fixed by physics or physiology. For example, the duration of vowel nasalisation before a nasal consonant varies with speech rate in American English, but not in Spanish (where it is roughly constant in duration), from which Solé infers that in Spanish, vowel nasalisation is a low-level coarticulatory effect due to the time required to lower the velum for the up-coming nasal consonant, whereas in American English vowel nasalisation 'vowel nasalization has become part of the language-specific instructions'. This and the various other cases in Solé's survey are extremely informative at teasing apart 'intentional' from 'mechanical' aspects of speech; though not quite synonymous with 'phonological' vs. 'phonetic', this is perhaps a more natural, less theory-dependent distinction.

Timing is also the focus of the six experiments presented by Manjari Ohala in 'Experimental methods in the study of Hindi geminate consonants'. In the most striking of these, Experiment 2, she finds that the consonant preceding the vowel that precedes a geminate is longer than in the corresponding non-geminate condition. That is, gemination in Hindi is a prosody, in this case a pattern of durational differences between words and without medial geminates that is evidenced throughout the initial syllable. The experiment was inspired by Local & Simpson's (1999) discovery of a similar non-local effect on the initial consonants of words containing geminates in Malayalam. But whereas Local & Simpson found shorter initial consonants with Malayalam medial geminates (perhaps suggesting a mechanism of temporal trade-off between the durations of successive vowels and consonants), the effect in Hindi goes in the other direction, suggesting a planned lengthening of the word-initial syllable.

The existence of (gradient) TRADING RELATIONS is one of the most significant discoveries concerning the relationship between phonological categories and observed phonetic forms. In 'Applying perceptual methods to the study of phonetic variation and sound change', Patrice Speeter Beddor, Anthony Brasher & Chandan Narayan present a perceptual experiment in which subjects identified tokens of *bet*, *bed*, *bent* and *bend* made with different proportions of vowel nasalization and nasal consonant duration, the two cues in some cases conflicting, in others cooperating. They found that some listeners treated vowel nasality and nasal consonant duration as perceptually equivalent, others did so only before /d/, and a third group gave more weight to vowel nasality than to nasal consonant duration. The pattern of trading relations and its variability between groups of listeners offers a promising explanation for the common historical change VN > Ṽ, as J. Ohala has proposed. M. Grazia Busà's chapter, 'Coarticulatory nasalization and phonological developments: Data from Italian and English nasal–fricative

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

sequences', takes this topic further, presenting aerometric data of oral vs. nasal timing in nasal + obstruent sequences in Northern vs. Central Italian and American English, finding (as expected) differences between all three. Northern Italian has a greater extent of anticipatory vowel nasalization, whereas Central Italian and American English a greater tendency to end nasality early, leading to a short stop closure between the nasal and following obstruent. Although both patterns are observed in both languages, it is the pattern with higher usage frequency that determines the direction of the sound change.

2.4 *Acoustic vs. articulatory categories*

Juliette Blevins's 'Interpreting misperception: Beauty is in the ear of the beholder' defends the importance of two aspects of John Ohala's mode of explanation of synchronic sound patterns: (i) misperception, which leads a language learner to think that a sound was produced in a different way than the speaker articulated it; and (ii) the non-teleological nature of sound changes, and the unnecessity of synchronic, structural markedness constraints. Misperception is well exemplified by nasal place assimilation to following oral consonants, and by the common sound change from dorsal [k^{j}] or [c] to coronal [ʧ], which Guion (1998) explains in terms of the acoustic similarity of [c] to [ʧ]. Blevins reports that Halle (2004) discounted Guion's acoustic and perceptual findings, preferring to explain the sound change solely in terms of articulatory feature geometry. Blevins remarks:

> In general, it seems that results of this sort encroach on territory which generativists are accustomed to viewing as part of the synchronic system. When this happens, there is resistance to the idea of incorporating experimental results. One consequence of this policy is an unrealistic conception of synchronic systems which, in many cases, duplicate explanations in other domains.

In contrast, Blevins's Evolutionary Phonology replaces synchronic teleology by diachronic phonetic developments which at each step might be motivated by perceptual, articulatory and/or lexical constraints. Synchronic variation is seen as offering a range of exemplars, allowing for different and sometimes ultimately contradictory patterns of development in different languages; for example, unstressed mid vowels may historically raise or lower in some languages (e.g. Bulgarian vs. Belorussian), yet in other cases unstressed high or low vowels may become mid, an apparent paradox presenting difficulties for a putative synchronic constraint relating stress to vowel quality. Blevins observes that the non-determinism of diachronic development sits naturally with stochastic views of phonology, such as that of Pierrehumbert (2002, 2003) or Keith Johnson's chapter.

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

Guion's acoustic explanation of the [c] >[ʧ] sound change (that they sound similar), cited with approbation by Blevins, is more convincing than the articulatory explanation (that their places of articulation are in close proximity) because the prevailing scheme of articulatory classification in phonological theory allots [c] and [ʧ] to quite distinct 'articulators', i.e. dorsal vs. coronal. Certainly, one might easily introduce a new articulatory category based on their common passive articulator, e.g. [palatal], but such a radical addition to current conceptions of articulatory geometry would be as conceptually 'costly' as simply restoring the acoustic feature [acute].

In 'A perceptual bridge between coronal and dorsal /r/', Olle Engstrand, Johan Frid & Björn Lindblom use acoustic measurements, articulatory modelling, acoustic waveform synthesis and perceptual testing to examine the historical development of coronal>dorsal /r/ (e.g. [r]>[ʁ] and their approximant congeners [ɹ] [ɥ̟], etc.) and the resultant synchronic variation common in many European languages. Through a detailed examination of their acoustics they show that despite being articulatorily quite distinct, coronal and dorsal variants lie on an acoustic continuum. Interestingly (though the authors do not comment on this), it is the frontest (alveolar) approximants, e.g. [ɹ̟] NOT the postalveolars or retroflexes, which are acoustically the most similar of the coronals to prevelar approximants [ɥ̟]. We have here positive evidence for restoration of the acoustic categories acute, i.e. [–grave], and [±flat]: in Jakobsonian terms both [ɹ̟] and [ɥ̟] are [–grave] and [–flat], as opposed to the [+grave] uvulars and pharyngeals, and to the [+flat] postalveolars and retroflexes.[4] Their categorical perception experiment shows that Swedish listeners discriminate rather poorly and non-categorically between coronal vs. dorsal /r/ (as expected, since they are not phonemically distinct in Swedish nor any known living language). The authors' sophisticated articulatory modelling yields a good match to the acoustic measurements for dorsal and retroflex approximants, but for alveolars the fit is poor, demonstrating that while articulatory models are interesting and important, they can be quite inaccurate in their acoustic output, compared to real speech. For this reason, we should be wary of over-enthusiasm about articulatory modelling.

2.5 *Abstract symbols vs. detailed examples*

During the last decade, exemplar-based phonology has taken root as a prominent alternative to generative phonology (including Optimality Theory). The exemplar-based approach is concerned with the cognitive grounding of phonological knowledge: although phonologists have long

[4] But note that contrary to Jakobson et al. 1952, pharyngeal approximants seem to be [–flat], with a relatively HIGH observed F3, and (according to Engstrand et al.'s articulatory modelling results) a RAISED F4.

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

held that phonological forms in the mental lexicon, as well as templates, constraints and rules, are to be expressed in an 'alphabet' of symbols that excludes the minutiae of specific instances, exemplar-based models envisage the storage of instances as they occur, with all their specific, individual details. New experiences are recognized as being similar to previous experiences by assessing the similarity of the new instance to the 'cloud' of stored exemplars. In essence, it is corpus-based phonology applied to the human mind. This approach has been assessed via various computational models, which are reviewed by Keith Johnson in 'Decisions and mechanisms in exemplar-based phonology'. It derives increasing support from experiments on human behaviour in speech perception and production, and has captured the interest of experimental phonologists partly because it offers mechanisms for explaining within-category variation in phonetic detail (thus also providing a foundation for sociophonetic variation) AS WELL AS a basis for abstraction of types and between-category differences. The key, exciting evidence comes from experiments that show: (a) the PLASTICITY of category boundaries, subject to various external factors (especially USAGE FREQUENCY differences between specific words, e.g. Pierrehumbert 2002, Wright 2003); Johnson shows (30–31) that vowels are reduced in different proportions in homophones with different usage frequencies; (b) listener sensitivity to WITHIN-CATEGORY differences (e.g. Beddor, Brasher & Narayan's chapter); and (c) listener sensitivity to FINE DETAILS – sometimes arbitrary or speaker-specific – which abstractionist approaches such as phonemics and generative phonology usually exclude as linguistically irrelevant (but cf. Terrance Nearey & Peter Assmann's chapter). For example, Goldinger 1996 discovered that lexical memory is sensitive to voice identity, words uttered by an unknown speaker (one of 10) on a single previous occasion a week earlier being remembered significantly more accurately than those produced by a new speaker, a result which flies in the face of the conventional expectation of phonologists (following Trubetzkoy 1939 [1969: 15–16]) that the details of a speaker's voice are 'abstracted away' by the listener, and form no part of the phonological representation of words. Hay, Warren & Drager 2006 discovered that in identifying diphthongs that are distinct in American English but nearly merged in New Zealand English, New Zealand listeners were biased by the experimenter's accent. In a subsequent experiment, Hay & Drager 2010 found subjects' vowel perception to be similarly biased by the national symbolism of stuffed toys (kiwis vs. kangaroos) that were MERELY PRESENT IN THE ROOM, the explanation being that the subject's category boundary is biased by the dialect primed by the toy, even though the subjects knew the speaker to be from New Zealand. This malleability of phonological category boundaries can be explained by reference to a pool of memories of prior experiences, but from an abstractionist view it is utterly unexpected and counterintuitive. The linkage of phonetic and non-phonetic – even non-linguistic – properties, such as visual and sociolinguistic properties of

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

experiences, has a top–down effect whereby expectations alter perception. As in generative phonology, there is no single, complete statement of 'Exemplar-based Phonology', but a variety of intersecting proposals (e.g. Blevins's 'Evolutionary Phonology' and Pierrehumbert's 'Stochastic Phonology') consistent with the central idea that specific experiences are memorized and used in recall, classification, etc. It seems that laboratory/experimental phonologists have at last begun to articulate a coherent and empirically preferable competitor to generative linguistics.

The new trend has its critics: in 'Probabilistic "sliding template" models for indirect vowel normalization', Terrance M. Nearey & Peter F. Assmann walk us through a succession of methods of normalization of speaker-specific variation in vowel acoustics, which they then competitively evaluate against data from three multi-speaker databases. To the extent that some of the models provide reasonably good vowel identification accuracy[5] and alternative accounts to some aspects of Johnson's model, this chapter offers a puissant defence of the abstractionist approach.

Eurie Shin's findings in 'How do listeners compensate for phonology?' are also claimed to support the generative view that (in this case) Korean listeners know the phonological rules that map underlying phonological forms onto observed speech signals, for example accepting intervocalic /s.p/ as a more likely source of [p.p] than /k.p/. The fact that /p.t/ or /k.t/ were sometimes considered by listeners to be possible sources of [t.t] (even though none of the [t.t] stimuli were assimilated forms) suggests to Shin that the subjects know and use the assimilation rules in processing the stimuli in these experiments – over-apply them, in fact. But this is not an argument to convince a doubter: for one thing, it disregards the bias caused by the tendency for subjects in a multiple-choice response format to use all of the available responses some of the time. Whether this could be an explanatory factor deserves investigation.

Two chapters seek to defend traditional generative phonology, while critically engaging with experimental phonology. In 'Morphophonemics and the lexicon: A case study from Turkish', Anne Pycha, Sharon Inkelas & Ronald Sprouse employ a corpus, the Turkish Electronic Living Lexicon, to study alternating (e.g. *kanat – kanada* 'wing' vs. 'wing.DAT') vs. non-alternating (e.g. *sanat – sanata* 'art' vs. 'art.DAT') words. They show that neither usage frequency, neighbourhood density, cohort size nor etymology effectively explains which forms alternate and which forms do not. But they then precipitously conclude that '[t]he more insightful analysis … thus appears to be the generative one, in which the underlying representation of each root contains the information necessary to predict its grammatical behavior' (385). This does not follow: in order to learn such underlying

[5] Rarely above 85% correct, however, or about one wrong vowel every two seconds.

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

representations and the rule of alternation, the Turkish learner needs to know the relevant word-pairs; in fact, simply memorizing them is both a necessary and sufficient basis for 'explaining' which stems alternate. There is no need for a learner to decompose the words into stem + suffix, abstract a single underlying stem form, infer the rule for the alternating forms and mark the non-alternating stems as resisting the rule, as the generative analysis would require, so by Occam's razor we should prefer simple memorization: the corpus/exemplar-based account proves stronger because it explains the same data without postulating unseen rules and underlying forms.

3. Some weaknesses of experimental or instrumental work

Let it not be thought that my advocacy for experimental approaches is blind to work which, though instrumental or mathematical, is nonetheless flawed. Neither theoreticians nor experimentalists have a monopoly on credibility or folly; we are all fallible, and unfortunately some of the chapters in this collection, however interesting to read, do not really convince. Perhaps these will offer some fertile ground for future arguments.

In Chapter 8, Jean-Luc Schwartz, Louis-Jean Boë & Christian Abry try to explain the phonological inventories of the UPSID database via a combination of three earlier models: Dispersion–Focalization theory (their variant of Liljencrants and Lindblom's Dispersion theory), J. Ohala's principle of 'Maximum Utilization of the Available Distinctive Features', and Perception-for-Action-Control theory, the authors' synthesis of auditory-acoustic and motor control models of speech. This is hardly reductionist: an awful lot of theory is invoked, making the chapter more like a survey of attempts to explain observed phonological systems in terms of the competition or interaction of production and perception constraints than a scientific theory. Despite one or two nice observations (e.g., they note that from the perspective of the Articulatory Dispersion theory, /y a ɯ/ should be as good a system of vowels as /i a u/), such a grand synthesis of individually reasonable proposals can hardly fail to be right in some parts; however, it is regrettably weak in its high ratio of model size to number of specific predictions.

In a technologically impressive but strikingly non-experimental Chapter 5, Jacqueline Vaissière aims 'to illustrate the usefulness of articulatory modelling as a tool for exploring the realisation of phonological contrasts' (54), a topic which though not new in phonetics (see e.g. Fant 1960, Stevens 1998) is given fresh relevance by focussing on the contrasts of French phonology, including the acoustics of front rounded vowels and nasal vowels. Despite similarities between some French vowels, English vowels and IPA Cardinal Vowels, Vaissière states that the acoustic target for French as opposed to English /i/ is 'most likely *highest F3* and not *highest F2*' (60), though she gives no direct evidence for this. The chapter is marred by ill-founded or

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

misconceived remarks concerning claimed shortcomings of the IPA. It is already well-established that American and British English /i/'s are pronounced a little differently (e.g. Ladefoged 2001: 207–209); that they are different from Cardinal [i] (IPA 1949: 20–21; Heffner 1950: 99), though French /i/ is said to be 'ABOUT cardinal' (IPA 1949: 21, my emphasis); that French /y/ and Swedish /y/ are different (Passy 1922: 101); and that different phoneticians' cardinal vowel productions are somewhat different. Such facts do not reflect 'embarrassing shortcomings' (61) of the IPA, the primary purposes of which are practical (originally pedagogical). The IPA DOES provide diacritics to distinguish some of these differences, if you wish to, and in so far as they are relevant to the phonology of a language, but it intentionally seeks not to capture speaker-specific and linguistically irrelevant details, so can hardly be faulted for this. The Cardinal Vowel system was specifically established not for French, English or Swedish, but as a language-independent basis for vowel notation (Abercrombie 1967: 155), a task for which it has proved remarkably resilient. The added sophistication provided by acoustic measurement and articulatory modelling gives no excuse to run down a useful and successful tool.

Rungpat Roengpitya examines Thai tones, spoken in a variety of contexts, in 'The variations, quantification, and generalizations of standard Thai tones'. Without explicit justification, her f_0 data is fitted to seventh-order polynomials. This seems like over-fitting: a seventh-order polynomial has three peaks and three troughs, properties not previously claimed for the tonal inventory of any language; see Grabe, Kochanski & Coleman (2007), who employ no more than a third-order fit, each order related to generally-established tonetic features such as average f_0, slope and convexity. Roengpitya identifies five tonetic processes: (i) end truncation, (ii) rate adjustment, (iii) range adjustment, (iv) plateau formation, and (v) phase realignment. These are not convincingly justified: phase realignment refers to the relative duration of a vowel + following nasal, and hardly affects f_0, suggesting that it is not tonetic. Range adjustment and truncation are not clearly independent: truncated tones have a reduced f_0 range; the illustration given of range adjustment in a falling-rising tone could be just as well interpreted as truncation of its falling part. Similarly, the illustrated range adjustment of a rising-falling tone might instead be interpreted as rate adjustment of the longer token. The examples of plateau insertion, which is only held to occur with contour tones, further support separation of falls from rises. The paper illustrates the fact that adopting quantitative methods is insufficient to prevent subjectivity from creeping into the methodology.

4. A RAPPROCHEMENT: CORPUS-BASED PHONOLOGY

Klaus J. Kohler's Chapter 4, 'Beyond laboratory phonology: The phonetics of speech communication', argues that large corpora of ecologically-natural

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

speech open up a new paradigm of experimental phonology, in which instead of taking phonological questions INTO the laboratory (letting phonological theorists delimit the material to be examined in the experimentalists' research), we can use experimental methods of e.g. acoustic analysis and statistics to examine very large quantities of naturally-occurring speech which has a genuine communicative function.

In 'Issues of phonological complexity: Statistical analysis of the relationship between syllable structures, segment inventories and tone contrasts', Ian Maddieson uses his phonological inventory database, UPSID (Maddieson 1984), to test the 'apparently widely held belief among linguists that complexity in one subpart of the grammar of a language is likely to be compensated for by simplification in another' (93). Maddieson finds support for one instance of this idea, that languages with many tones tend to have simpler syllable structure; but he disproves the hypothesis that a language with many consonants should have few vowels. Though this is true in individual cases (e.g. Abkhaz), it is not generally true, as previously noted by Stephens & Justeson (1984).

5. EXPERIMENTS VS. OTHER METHODS

In the preceding sections, I have detailed numerous cases in which the results of experimental phonology either add to or, in some cases, seriously challenge some proposals from theoretical (mainly generative) phonology. Nevertheless, we have also seen that experimental and/or quantitative work is not uniformly 'good', either. To conclude, let us briefly consider what makes the rigour of the experimental method work, and where other methods may nevertheless lead to work of value.

John Ohala's Berkeley colleague, the phonologist Larry Hyman, argues in 'Elicitation as experimental phonology' that the standard procedures of informant-based phonology constitute experiments, illustrating the point through the development of an analysis of some tonological data from Thlangtlang Lai, a Kuki-Chin language of Burma. Certainly, there is a process of data collection, hypothesis formation, and further data collection with the intention of refuting or supporting the hypotheses. True, elicitation is a part of many experiments, but EXPERIMENTAL elicition needs to be done with properly prepared stimuli and a properly defined task for all subjects. Hyman offers a non-dogmatic exploration of alternative possibilities, and one appreciates that he is attempting, in good humour, to bridge the divide for the benefit of the honorand and the occasion.

But even though the notebook method is based on observations, no, these are not experiments. There are no physical recordings, without which an informant's speech events are unavailable for public scrutiny or for the researcher's reexamination: an important part of data verification. Where are the measurements? Phonetic transcriptions are not records, but

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

are themselves 'pre-cooked' theory-laden interpretations of the transcriber's auditory impressions. Since none of us is perfect, they almost certainly contain a proportion of errors (regrettably unquantifiable, without permanent records and quantitative measurements) and they are subject to 'experimenter' bias: Hay & Drager 2010 demonstrate that such biases may be far more subtle than anyone has previously dreamed. Though we know that language is riddled with variability, there is typically a statistically insufficient number of subjects, little or no reported variation, no statistics and hence little SCIENTIFIC credibility in traditional phonological methods. This uncomfortable fact partly explains why colleagues in other disciplines, such as psychology, neuroscience, and physiology, find much of linguistics hard to swallow, to the point of ignoring, marginalising or sometimes even deriding it, preferring instead to undertake experimental studies of language outside the confines of linguistics. Linguistics has no monopoly claim to being 'the scientific study of language', as we may pitch it to students; for the most part linguistics is not a scientific study at all. With this in mind, perhaps we should urge that the proper places for impressionistic, notebook research are (a) pre-experimental notes of 'observations to be followed up in more detail later', or (b) in cases where access to speakers may be difficult (as in documentation of endangered languages), transcription is an 'intermediate technology' for rapid – though possibly error-prone – recording.

So let us not pretend that notebook linguistics is experimental; Hyman's simile is light-hearted. Though we may look to phonological theory as a source of interesting hypotheses, proposals, and semi-formal observations of linguistic phenomena, we should accept the simile as a declaration of goodwill towards experimental linguistics, in the expectation that as experimental methods become widely adopted in linguistics, the subjectivity of notebook-based impressionistic research may be replaced by a firmer understanding of language as a natural phenomenon, to be studied like any other aspect of nature.

REFERENCES

Abercrombie, David. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Bonatti, Luca L., Marcela Peña, Marina Nespor & Jacques Mehler. 2005. Linguistic constraints on statistical computations. *Psychological Science* 16.6, 451–459.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Caramazza, Alfonso, Doriana Chialant, Rita Capasso & Gabriele Miceli. 2000. Separable processing of consonants and vowels. *Nature* 403, 428–430.

Cercle Linguistique de Prague. 1931. *Réunion Phonologique Internationale Tenue à Prague (18–21/XII 1930)* (Travaux du Cercle Linguistique de Prague 4).

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Chung, Sandra. 1983. Transderivational relationships in Chamorro phonology. *Language* 59.1, 35–66.

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

Coleman, John. 2003. Discovering the acoustic correlates of phonological contrasts. *Journal of Phonetics* 31, 351–372.
Cooper, Franklin S., Pierre C. Delattre, Alvin M. Liberman, John M. Borst & Louis J. Gerstman. 1952. Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America* 24.6, 597–606.
Damasio, Antonio R. & Daniel Tranel. 1993. Nouns and verbs are retrieved with differently distributed neural systems. *Proceedings of the National Academy of Sciences of the USA* 90, 4957–4960.
Eichhoff, Frédéric Gustave. 1836. *Parallèle des langues de l'Europe et de l'Inde*. Paris: Imprimerie Royale.
Fant, C. Gunnar M. 1960. *Acoustic theory of speech production*. The Hague: Mouton.
Godfrey, John J., Edward C. Holliman & Jane McDaniel. 1992. SWITCHBOARD: Telephone Speech Corpus for Research and Development. *ICASSP-92, IEEE International Conference on Acoustics, Speech and Signal Processing, 1992*, vol. I, 517–520.
Goldinger, Stephen D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22, 1166–1183.
Goldsmith, John. 2001. Unsupervised learning of the morphology of a natural language. *Computational Linguistics* 27.2, 153–198.
Grabe, Esther, Greg Kochanski & John Coleman. 2007. Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and Speech* 50.3, 281–310.
Guion, Susan G. 1998. The role of perception in the sound change of velar palatalization. *Phonetica* 55, 18–52.
Halle, Morris. 1959. *The sound pattern of Russian: A linguistic and acoustical investigation*. The Hague: Mouton.
Halle, Morris. 2004. Moving on. Plenary address in honor of the 80th anniversary of the founding of the Linguistic Society of America, Boston, MA, 8–11 January 2004.
Halle, Morris & Kenneth N. Stevens. 1964. Speech recognition: A model and a program for research. In Jerry A. Fodor & Jerrold J. Katz (eds.), *The structure of language: Readings in the philosophy of language*, 604–612. Englewood Cliffs, NJ: Prentice-Hall.
Hay, Jennifer & Katie Drager. 2010. Stuffed toys and speech perception. *Linguistics* 48.4, 865–892.
Hay, Jennifer, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34, 458–484.
Heffner, Roe-Merrill Secrist. 1950. *General phonetics*. Madison, WI: The University of Wisconsin Press.
IPA [International Phonetic Association]. 1949. *The principles of the International Phonetic Association*. London: International Phonetic Association.
Jakobson, Roman. 1939. Observations sur le classement phonologique des consonnes. *The Third International Congress of Phonetic Sciences*, Ghent. [Reprinted in Roman Jakobson, *Selected writings*, vol. 1, 272–279; Berlin: Mouton de Gruyter, 1982.]
Jakobson, Roman. 1962. The phonemic concept of distinctive features. In Antti Sovijärvi & Pentti Aalto (eds.), *The Fourth International Congress of Phonetic Sciences* (ICPhS 4), 441–455. The Hague: Mouton.
Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1951. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: The MIT Press.
Just, Marcel Adam, Patricia A. Carpenter, Timothy A. Keller, William F. Eddy & Keith R. Thulborn. 1996. Brain activation modulated by sentence comprehension. *Science* 274, 114–116.
Kempelen, Wolfgang von. 1791. *Le mécanisme de la parole, suivi de la description d'une machine parlante*. Vienna.
Klagstad, Harold L. 1958. The phonemic system of colloquial standard Bulgarian. *The Slavic and East European Journal* 2.1, 42–54.
Ladefoged, Peter. 2001. *A course in phonetics*, 4th edn. Fort Worth, TX: Harcourt College.
Liberman, Alvin M. 1957. Some results of research on speech perception. *Journal of the Acoustical Society of America* 29.1, 117–123.
Liberman, Alvin M., Franklin S. Cooper, D. P. Shankweiler & Michael Studdert-Kennedy. 1967. Perception of the speech code. *Psychological Review* 74.6, 431–461.

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press

Local, John & Adrian Simpson. 1999. Phonetic implementation of geminates in Malayalam nouns. In John J. Ohala, Yoko Hasegawa, Manjari Ohala & Daniel Granville (eds.), *XIVth International Congress of Phonetic Sciences* (ICPhS 14), vol. 1, 595–598. Berkeley: University of California, Department of Linguistics.
Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.
Passy, Paul. 1922. *Petite phonétique comparée des principales langues européenes*, 3rd edn. Leipzig: Teubner.
Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory phonology 7*, 101–140. Berlin: Mouton de Gruyter.
Pierrehumbert, Janet B. 2003. Probabilistic phonology: Discrimination and robustness. In Rens Bod, Jennifer Hay & Stefanie Jannedy (eds.), *Probabilistic linguistics*, 177–228. Cambridge, MA: The MIT Press.
Rapp, Reinhard. 2007. Part-of-speech discovery by clustering contextual features. In Reinhold Decker & Hans-J. Lenz (eds.), *Advances in data analysis: Proceedings of the 30th Annual Conference of the Gesellschaft für Klassifikation e.V., Frei Universität Berlin, March 8–10, 2006*, 627–634. Berlin: Springer.
Saffran, Jenny R., Richard N. Aslin & Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.
Stephens, Laurence D. & John S. Justeson. 1984. On the relationship between numbers of vowels and consonants in phonological systems. *Linguistics* 22, 531–545.
Stevens, Kenneth N. 1998. *Acoustic phonetics*. Cambridge, MA: The MIT Press.
Stevens, Kenneth N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America* 111.4, 1872–1891.
Trubetzkoy, Nikolai Sergeevitch. 1939. *Grundzüge der Phonologie*. [Page references are to the 1969 English version, *Principles of phonology*, translated by Christiane A. M. Baltaxe, University of California Press.]
Wright, Richard. 2003. Factors of lexical competition in vowel articulation. In John Local, Richard Ogden & Rosalind Temple (eds.), *Phonetic interpretation: Papers in laboratory phonology IV*, 75–87. Cambridge: Cambridge University Press.

Author's address: Phonetics Laboratory, University of Oxford, 41 Wellington Square, Oxford OX1 2JF, UK
john.coleman@phon.ox.ac.uk

https://doi.org/10.1017/S002222671000040X Published online by Cambridge University Press