A Consumer Guide to Phonological Evidence

Marc van Oostendorp
Meertens Instituut

Abstract

Modern phonological theory is confronted with a wealth of new data from many different sources. This paper gives a summary and taxonomy of the kinds of evidence we currently have at our disposal. For each type it briefly discusses how it has been used, and gives some of the advantages and disadvantages. The main distinction made is between 'existing' data and 'invented' data, even though it is shown that such distinctions should be considered very carefully. I argue that each type of data can have a function in phonological argumentation, but also that each type of data has some very specific problems. The best evidence is converging evidence from different sources.

1. Introduction

Data that might be useful for the phonologist can come from many different angles. The edifice of phonological theory is built on a foundation of minimal pair tests, phonetic experiments, philology, extensive fieldwork and many other methodologies.[1]

On the one hand, this multitude of methodologies is healthy. If psycholinguistic experimentation points in the same direction as careful reasoning about the metrics of classical poetry, the different types of evidence might support each other.

At the same time, there are many worries. A lot of data on which we build our theories are contested, or contestable. They seem to be based on unreliable sources, or unsuitable methodologies, or they just seem too irrelevant for the question that is being asked.

There is hardly any discussion among phonologists about these issues. The little discussion there is often takes a polemical form, arguing that certain types of evidence are completely unsuitable for a discipline which aims to call itself scientific, or that other types of data inherently are not about phonology, but about something else.

This is not the point of departure I will take here. The basic assumption underlying this guide is that (synchronic) phonology is about (a speaker's) knowledge of sound systems.[2] Direct evidence about this knowledge does not exist, and may never exist. All our evidence is therefore by necessity indirect, and has to be handled with some care. But for the same reason, it is very difficult to dismiss any *type* of data out of hand. Any product of linguistic activity potentially sheds light on the knowledge which human beings have of the sound structure of their language.

This guide has been written with the phonological theorist in mind, who wants to make sense of all the data we have at our disposal nowadays. It is not my goal to explain how different methodologies should be applied to obtain certain types of data. It is my goal to consider how the data which are thus acquired can be used in constructing and testing phonological theories. The reader should keep in mind that the usefulness of empirical data depends to a large extent on the theoretical goals one has, and I try to pay attention to this issue as well.

[1] The title of this paper obviously is a tribute to Ohala (1986). The paper furthermore pays homage to Tromsø thinking about methodological thinking.

[2] There are other views of phonology of course; the most prominent alternative holds that the patterns of phonology are caused not so much by the structure of the mind as by the properties of human societies and the way in which people interact with each other. Taking such a stance, which is perfectly reasonable, will probably change to some extent the considerations followed here. Many of the data sources cited here will still be useful for the phonologist who follows this line of argument; but also in this case, there is nothing like 'internal' evidence, since we cannot look into the common language structure of a society directly. I choose the stance that language structure is inherent to, caused by, and reflective of the human mind simply because it is the one I am most familiar with, and it is is the one that has underlied the past 10 years of CASTL Phonology, as well as the thinking of Curt Rice.

© 2013 Marc van Oostendorp. *Nordlyd* 40.1: 274-293, special issue 'A Festschrift on the Occasion of X Years of CASTL Phonology and Curt Rice's Lth Birthday', ed. by Sylvia Blaho, Martin Krämer and Bruce Morén-Duolljá. University of Tromsø.
http://www.ub.uit.no/baser/nordlyd

de Lacy (2009) has rightfully pointed out that evidence does not exist outside of theories. de Lacy (2009)'s own paper provides a very insightful discussion of phonological evidence from a slightly different angle; he discusses which kinds of data provide evidence for a specific theory of phonological knowledge, identified by him as 'generative, innatist and modular' (GIM). Although this is a perfectly valid approach, it means that certain types of evidence will fall outside of the scope of the theory — as a matter of fact that is the goal of the whole exercise. For instance, 'frequency effects' fall outside of the scope of the theory since the GIM framework says that such effects, if they exist at all, have to be accounted for by some outside scientific discipline. Again, this is a perfectly valid line of reasoning, but here we just make the scope somewhat wider: given that we assume that phonological theory is about knowledge, what constitutes potential evidence as to the nature of that knowledge?

A different way to explain it is the following. A GIM framework as sketched by de Lacy (2009) will typically assume that there is a 'phonetic component' elsewhere in the cognitive make-up of the speaker, and responsible for 'phonetic implementation'. It will relegate certain types of evidence, e.g. with respect to loanword adaptation to this other module. The approach taken here will make no distinction between GIM and phonetic implementation; not because it denies that such a distinction exists, but because such a distinction is too finegrained for our goal here. On the other hand if some aspect of phonetic transmission is purely physical (e.g. based on acoustic properties of the signal) or physiological, it does not constitute evidence from our point of view either.

A third way to see the difference is that Prince (2007) distinguishes between 'Free-Standing Theories' (FST)— theories as formal objects of which the formal properties can be explored — and 'Theories of Data' (TOD) — 'which produce analyses when set to work on collections of facts'. Prince (2007) clearly prefers the former type of theory and de Lacy (2009) follows him in this. I do not want to go into the methodological debate as to the comparison of these two views on theory (it seems to me that in a healthy discipline, one needs both), but in any case the emphasis in this paper will be one of TOD; see 6 for some discussion of FST criteria.

The taxonomy of types of evidence I present here, mostly serves a pedagogical function and should not be taken too seriously otherwise. However, linguists sometimes refer to another taxonomy: the difference between 'internal' and 'external' evidence (Zwicky 1980). The term derives from literary criticism (Wimsatt and Beardsley 1946) where it refers to arguments derived from a literary text from arguments from other sources, such as the biography of the author. In the early days of generative grammar, there were heated debates about the relevance of 'external' evidence; nowadays types of 'internal' evidence seem more the target of attack. It seems to me that there is no real point; I fail to see what would constitute 'internal' evidence for knowledge. We know for instance that native speaker judgements are susceptible to a process of self-monitoring which may be partly based on psychological processes different from the grammar (e.g. prescriptive and other biases, cf. Boersma 2004, Keller and Asudeh 2002, Labov 1977; 1987; 1996, Schütze 1996:for a few views), so evidence which is based on such a methodology is not necessarily purely linguistic. It seems to me that therefore the dichotomy between internal and external evidence is unusable; also because it seems to reflect more the methodology that is being used (introspection vs. other means of study) than the actual evidence that is the result of the methodology.

However, it is still reflected to some extent in this guide. The evidence in sections 2 (Classical Evidence) and 6 (Formal Evidence) deals mostly with types of 'internal evidence'; the sections 4 (Artificial Evidence), 3 (Massive Evidence) and 5 (Experimental Evidence) with evidence that is obtained mostly by 'external' methodology.

This guide does not deal with data collecting methodology directly, although at some points it is unavoidable to point out some risks in collecting data of one type or another. But it does deal with methodology of a different type, since being able to evaluate different types of data carefully is one of the theoretical phonologists' skills.

2. Existing Data 1: Classical Evidence

'Classical' evidence is a name for the data sets that have a long tradition in (synchronic) phonology, tracing its roots in structuralist analysis. Data are often acquired by introspection, either of the linguist himself or of some informant, but this is not necessarily the case. All the types of evidence discussed here could equally be extracted from a corpus, for instance.

2.1. Minimal pairs

- **Applicable to:** Analysis of contrast
- **Example:** From the fact that *man* and *ban* are distinct words we can conclude that nasality is contrastive in English.
- **Pro:** The minimal pair test usually leads to clear and intersubjective data
- **Contra:** (i) problems with phonemic analysis (e.g. the famous /ŋ - h/ problem), (ii) what exactly is a 'minimal pair'?, (iii) often it is just not possible to find minimal pairs for all sounds due to lexical sparsity.

This undoubtedly is the most classical type of genuinely phonological data. It stems from structuralist work. It is used to establish phonological contrast, and it continues to do a rather good job.

This does not mean that many problems have not been identified with it over the years. Well-known is the problem that a minimal pair test leads to the result that [h] and [ŋ] are allophones: given their distribution ([h] only occurs at the beginning of feet and [ŋ] only in the coda of syllables) no minimal pair can be constructed; yet, in some analyses it would still be necessary to consider them as different segments, for instance because of their phonetic dissimilarity.

Another problem is that contrast can sometimes shift from one segment to another. A famous problem, raised originally by Joos (1942), is that of *Canadian Raising*. Joos points out that the words *writer* and *rider* sound as [rʌyrər] and [rayrər], respectively. This is a minimal pair, at least according to some definitions (namely those which allow the forms in the minimal pair to be derived), and it could lead us to posit two different diphthongs for Canadian English (that was in fact Joos (1942)'s proposal).

However, it has also been argued that the distribution of the diphthongs is predictable, since it is dependent on the underlying quality of the following consonant, which is voiceless for [rʌyrər] and voiced for [rayrər]; the contrast between the consonants is no longer visible on the surface since a process of flapping has applied. We thus have two options: we could either decide that only the underlying contrast counts, in which case we should not look at surface minimal pairs, but at underlying minimal pairs instead (/rayt+ər/ and /rayd+ər/) or alternatively only on underived forms. This position was famously defended by Chomsky (1957; 1964) in the early days of generative grammar, and still has its defenders (e.g. Idsardi 2006). Other phonologists however, prefer Joos' view, for instance because using it would involve the notion of opacity, which involves extrinsic rule ordering, that is disallowed in several frameworks.

2.2. *Distribution*

- **Applicable to:** Static (often phonotactic) evidence in the lexicon
- **Example:** English has no words starting with /rt-/
- **Pro:** Clear intersubjective patterns
- **Contra:** (Absence of) patterns could be result of historic accident (the fact that English does not have such words follows from the fact that they were never there in any source language and there was no development leading to them); what do we do about 'exceptional patterns' (e.g. learned words); distributional evidence does not show productive phonological activity.

Distributional evidence is another classic, used already in structuralist phonology to describe phonotactics. From some source — an existing word list or dictionary, a corpus, the memory of a native speaker — we observe which combinations of elements are possible and which are not. Very often, this concerns phonotactics: words start with rising sonority, ŋ is not allowed at the beginning of the word, etc., but also the internal structure of segments can be studied in this way: in Dutch, we find the phonemes /p, t, k, b, d/ but not */g/, hence the combination of [velar] and [voice] is disallowed.

On the one hand, most of the generalisations we obtain from distributional evidence is very clear. However, it is not always necessarily clear that such static patterns are part of the knowledge of a native speaker. The lexicon of a given language is shaped at least in part by historical accident. For instance, the fact that /ŋ/ in English cannot be word initial can be understood as a combination of (i) the fact that /ŋ/ historically derives from /nɡ/ sequences, and nasal - plosive sequences generally were not or are not allowed in English, and (ii) there are no serious borrowings of words starting with /ŋ/ from other languages. Thus from the fact that in our distribution we find no /ŋ/ at the beginning of words, it does not necessarily follow that there is a constraint or a synchronic rule or anything of the sort disallowing such a word-initial segment.

An extra complication arises in cases in which there seem to be exceptions. For instance, one could say that English does not have words starting with /tl-/ either. Now the online Merriam-Webster dictionary has three items starting with this sequence: *Tlaxcala, Tlingit* and *Tlemcen*. Obviously, these are all three foreign geographical names; they are therefore sometimes discarded as pieces of evidence for phonological analysis, but one has to be careful there. For instance, Merriam-Webster gives pronunciation information for these names, and in each case it notes /tl/ at the beginning, whereas for names starting with *ng*, it notes some phonological change (*Ngauruhoe* /əŋ-ɡau-rə-hō-ē/). But even if we accept names as potential evidence, we might be tempted to say that their number is too small (there are thousands of words with *tr, pr, kr, kl* and *pl*, and only three with *tl*). This cannot be coincidental, and one might feel tempted to say that for instance the Obligatory Contour Principle (OCP) is responsible for the effect.

An interesting paper in this respect is Becker et al. (2011), who point out that certain statistical tendencies in the Turkish lexicon correspond to native speaker's behaviour in wug tests, whereas others do not. Furthermore, they argue that there is a correlation between the phonetic 'naturalness' of a pattern and the readiness of subjects in an experiment to abide by the pattern when encountering new forms.

In connection to *Tlingit* etc., I noted that names are sometimes considered to be exceptional. Something similar is true for at least interjections and numerals. Furthermore, distinctions in distributional patterns are often made between stems and affixes and betwen function words and lexical words, or even between certain classes of lexical words (e.g. English noun-verb pairs such as *récord-recórd*. This means that the distributional patterns observed for one class are not always considered relevant for the other classes (Smith 2001; 2011).

2.3. Phonological alternations

- **Applicable to:** Phonological processes (rules, etc.) of any sort
- **Example:** Final Devoicing in Dutch: hɔnt-hɔndə
- **Pro:** Gives a window on phonological activity, at least if productive.
- **Contra:** It is not necessarily clear how we deal with processes that are not, or only marginally productive (English velar softening in *electricity*)? How do we know that the individual forms are not stored in the lexicon?

While the previous types of evidence both are useful for describing static properties of a language — its inventories of phonemes, syllables, and other constituents — the classical type of evidence for 'active' phonology — rules, processes, or however one wants to call them — are phonological alternations: a morpheme shows up differently in one context than in another.

The important criterion for the validity of a phonological alternation as a piece of evidence is that it be 'productive', which in turn means that it should have no lexical exceptions, or, in other words, it should apply whenever its condition is met (Kaye 1995).

The interpretation of this criterion is open for discussion. In several traditions of phonology, there is place for concepts such as 'minor rules' (see Zonneveld 1978:for discussion) or 'cophonologies', applying only to specific subsets of phonological data (Inkelas et al. 1996, Simon and Wiese 2011). In such theories, also non-productive alternations can find a place as evidence. The question in that case is how far we are willing to go in granting processes phonological status. For instance, allowing 'velar softening' (the alternation in *elektri*[k] - *elektri*[s]*ity*) into our phonological theory might severely complicate our theories of what is a phonological representation or what is a phonological change. More importantly, the fact that two words are obviously related in form and meaning does not necessarily lead to the conclusion that they are one and the same morpheme (Lightner (1972)'s suggestions that *teeth* and *dentist* might be related in phonology does not seem to have many followers any more).

On the other hand, absence of an alternation is not always taken as evidence for absence of a corresponding process, since the process may be blocked — obviously depending on one's theory — by additional phonological restrictions, by morphological boundaries, in non-derived environments, opaque rule ordering, etc. (Bakovic 2011, Burzio 2011:to cite a few references).

Hyman (2007) argues that fieldwork methods like those resulting in phonological alternations should be considered 'experimental'. He points out that in obtaining data about tonal alternations in Kuki-Chin languages he had to ask informants about rather unnatural sequences such as the following (I leave out the actual words, since they are irrelevant):

(1) a. 'pig's friend's graves price'
b. 'chief's beetle's kidney basket'
c. 'monkey's enemy's snake's ear'

"It wouldnt impress any psychologist," Hyman writes, "and it would definitely horrify an anthropologist, but who else but an experimental linguist could present such stimuli?" He also states that it "is most significant that the novel utterances are produced with the appropriate application of tone rules." It should be pointed out, however, that such methodologies can be most succesfully applied in cases like this one in which productive external sandhi is applied. For this reason, this methodology is dominant in syntactic theory. (For interesting views on problems in the methodology, see Hay and Drager (2010).)

2.4. Descriptive grammars

- **Applicable to:** Anything phonological that happens to be described
- **Example:** Yawelmani used to have a rule deleting consonants between two vowels (Yawelmani is actually extinct).
- **Pro:** This is a way of finding many interesting patterns at relatively low cost — no fieldwork needs to be done. It is particularly useful as a first impression.
- **Contra:** Data are seldom complete with respect to the question you are asking; reliability of the author is always a problem (also in the sense of the author having probably worked from a completely different point of view).

A lot of phonological literature is based on data that have been taken from descriptive grammars, grammatical sketches, word lists made by fieldworkers, etc. Especially typologically oriented work would be difficult to do without being able to rely on these sources. All scholarly work is based on trust — relying on the belief that other people have done a proper job and the data are therefore reliable —, and this is true in particular for theorists, who work with other people's data to begin with. In a field like phonology, in which evidence can come from so many sides, so the theorist cannot be expected to be well versed in all relevant methodologies.

The data in descriptive grammars have usually been built up from the kinds of evidence that has been described in the preceding sections. They therefore suffer from similar problems, and often in a more serious form, since authors do not describe their methodology in obtaining the data; the transcriptions may be more or less impressionistic, for instance. Similarly, descriptive grammars are by necessity limited in size and may not always provide the theorist exactly the data she needs in order to confirm or refute her theory. One should be extremely careful in drawing hard conclusions from such a description, in particular when it is about complex points.

In typological work, the problems may diminish since methodological and other deviances of individual fieldworkers may be statistically filtered out from the data: we can than possibly hope that the mistakes and biases of one fieldworker will be balanced by mistakes and biases by another fieldworker in the opposite direction. (Although, as one reviewer points out, most fieldworkers will be speakers of a small set of Indo-European languages and therefore share certain biases.)

3. Existing Data 2: Massive Evidence

Modern information and computing technologies allow us to set a step beyond the 'classical' evidence. In this section, I discuss the kinds of evidence which come from larger sets of data (or partly new sources of data) than were available a few decades ago. I call these 'massive' evidence — of course, in English this suggests that the evidence is particularly compelling, but in this guide I prefer to stay neutral on that isse.

3.1. Corpus frequency

- **Applicable to:** Mostly static patterns, and 'markedness', but also productivity of morphonological processes.
- **Example:** *t* is the most frequent plosive in English, whereas *k* is in Japanese (Yoneyama et al. 2003, Beckman et al. 2003).
- **Pro:** Objective, easy to check for other researchers, gives more fine-grained distinctions.
- **Contra:** the relevance of frequency for phonology is not uncontested; corpus distribution can be influenced by many factors that are potentially irrelevant; most importantly, corpora tend to not contain phenomena which are marginal and might be of interest exactly for that reason; the reasons why some forms might be absent in a (necessarily finite) corpus might be immaterial to the 'real' phonological pattern.

The first type of evidence concerns the relative frequency in a corpus. This type of evidence has played an informal role in phonological argumentation for a long time. Many papers of the last few decades contain footnotes which acknowledge that there are some counterexamples to a generalisation, but these are 'only a handful'. Furthermore, markedness theory often has a statistical flavour Hume (To appear): the fact that segment x is more frequent than segment y in a language, is one of the operational definitions of markedness.

A large advantage of the present-day availability of large phonological corpora is that claims such as these can be easily tested. Furthermore, the results of a corpus search are usually relatively easily replicable, at least if the corpora are publicly available. And finally, the quantitative nature of these data make it possible to do much more fine-grained statistical analysis than typical grammaticality judgements would. (See Durand et al. 2012:for an overview of available corpora for phonological research as well as discussion of their use.)

Yet the relevance of corpus frequency data is all but uncontested. According to a (still) fairly standard view, phonology, being a module of the grammar, deals only with categorical data — everything gradient is the domain of phonetics. On the other hand, even if one takes this position, this does not necessarily mean that one does not make at least some predictions on gradual differences in grammaticality.

Furthermore, even if one subscribes to a categorical view of phonology, giving only binary results – certain things are grammatical, other things are ungrammatical — one could still acknowledge that the actual data in the world are influenced by many factors which are not grammatical — Chomsky (1965)'s famous 'performance' factors. Given the number and complexity of such factors, the grammar might be expected to only have a statistical effect on the data.

There is also a disadvantage of corpus frequency data connected to this. The distribution of data in a corpus is influenced by many factors, which are not always under one's control. As far as I am aware, there is at present no golden standard of word frequency in English: I do not think there is an uncontested Top-10 of the most frequent words in this language. The reason for this is that results differ (slightly) from one corpus to the next, depending on whether it is a written or spoken corpus, whether the material is edited, how old it is, etc.

Finally, corpus data are definitely not suitable for all phonological questions. For instance, very marginal contrasts — which are too small to play a role in certain statistical analyses — can still be of interest to the phonologist.

3.2. Typology

- **Applicable to:** Anything phonological
- **Example:** No language has more than 200 phonemes.
- **Pro:** Is basically our only window into the validity of claims on (implicational) universals.
- **Contra:** Data are often based on descriptive grammars and share the disadvantages of those; the set of existing languages (let alone that of described languages) may not be a good sample of the set of possible languages.

Typological databases are for cross-linguistic comparison what corpora are for the study of the sound system of an individual language: a way to test empirical claims that is neutral, and that can be replicated if the databases are available.

Since there are very few known absolute universals with any interesting emperical content, most work on universals in phonology concentrates on implicational universals (of the type 'if a language has property P_1, it should also have property P_2'). Such universals can only really be tested with evidence from general typological databases (such as the WALS) as well as more specialized phonological databases (such as UPSID and P Base for phonological inventories, and StressTyp for stress). Typologists have in recent years put a lot of effort in building balanced databases of languages of the world; these can be very useful for many kinds of phonological analysis.

An important reason to be very careful when dealing with typological databases is that the data are usually gathered from existing descriptive grammars (section 2.4) or other such sources. The problems connected to such sources will also exist for databases of this type, and may in some cases be aggravated by the fact that there have been many different authors involved, and it is difficult to verify all of their methodologies. If there is a lot of data, the 'noise' which such factors bring, might be factored out in the statistics, but it is wise to always remain aware of such problems.

There has been quite some discussion in recent years on the use of gaps in the typology as evidence for phonological theory: if it is not likely that the set of existing (or sufficiently known and documented) languages is representative for the superset of all languages in the world, we cannot use the absence of a predicted language ('overgeneration') as an argument against a theory, at least according to Hale and Reiss (2008). Even when we try to do a statistic analysis, the problem is that we have no idea how big the set of possible languages is, and therefore how the existing set relates to that superset.

It seems to me that the problem with this argument is that we will always need some measure against overgeneration, since otherwise we cannot claim to have a theory of phonology. A Turing Machine can generate any language imaginable including natural languages, so if we have a Turing Machine and overgeneration is not an issue, we can propose the Turing Machine as our theory and have solved the problem of universal phonology, although we would of course still have to solve the issue how to describe the phonologies of individual languages.

One could also argue that restrictiveness is simply a value, even in the context of small sample size, since it makes our theories more falsifiable at least in theory. A theory disallowing something which does not exist is inherently better off than a theory which allows the same phenomenon.

3.3. Variation data

- **Applicable to:** Anything (within a language).
- **Example:** Dutch dialects do not differ systematically with respect to placement of stress within monomorphemes, but they do differ with respect to compound stress.
- **Pro:** Can give a much more fine-grained view in what (micro)parametric variation looks like.
- **Contra:** Same as for typology (although the data are sometimes more under control)

Datbases on language variation (e.g. dialectological databases or collections of data about different sociolects) are typological databases, but with variation on a smaller scale. They can serve similar functions, for instance, checking the validity of implicational relationships within varieties of a language or a language family. Obviously, it shares the same advantages and the same problems, although there are shades of differences. The data in dialectological databases tend to be gathered in a more systematic and coherent way across different datapoints — data in large typological databases have often been gathered by researchers with very different concerns and backgrounds; this may be less the case in dialectological databases —, so that the different datapoints can be compared more easily.

It is sometimes pointed out that the advantage of studying microvariation over variation at a larger scale is that it allows us to determine more precisely the locus of variation parameters (Barbiers 2009): if we compare, say, Japanese to Swahili, there are so many differences that it is difficult to see which differences correlate with which other. In comparing different Japanese dialects to each other, on the other hand, we keep many things constant, so that we can more easily isolate certain points of variation from each other.

By the way, even though I put variation data in this section on 'massive data', sometimes small amount of data (comparing 3 or 4 dialects on 1 or 2 points of variation) can already provide a lot of insight.

3.4. Language change

- **Applicable to:** Mostly distributional facts
- **Example:** Compensatory lengthening in the history of English (*gans* > *goose*) (Kavitskaya 2002)
- **Pro:** Language change is arguably constrained by phonological factors.
- **Contra:** Language change is also subject to many other factors which are very difficult to control for; some of these factors may even be completely unknown.

Like language variation, also language change data do not necessarily have to be massive, or collected in digital databases, in order to be useful, although more and more data become available in this format, and one should use them whenever possible.

The issue is to what extent historical changes provide real evidence on synchronic phonology. It is not clear that the set of possible historical changes is exactly the same as the set of possible productive synchronic alternations, although it is of course the case that many synchronic alternations (as well as all of language variation) have their origin in diachronic change, or are at least mirrored by it. Also, it is obvious

that language change is constrained by the synchronic phonology: by definition, language change can never lead to a synchronically impossible language (also all intermediate stages in a change have to be possible languages).

Actually, there are also demonstrably synchronic alternations which are not possible (or at least likely) diachronic changes. There are examples of so-called 'rule telescoping' (Bach and Harms 1972, Hyman 1975): a series of inherently natural phonological changes leads to an alternation which is unnatural. For instance, in Tswana, a series of changes, partially still visible in dialects, led to an alternation between a labial and a palatal (Bateman 2007, Kochetov 2011). It is probably impossible that such a change would ever happen in a language directly:

(2) a. *Tswana*: /-ga-wa/ → [gatʃwa] 'request (PASS)'
b. *Northern Sotho*: [gapʃa] (labialization moves from fricative to plosive)
c. *Lobedu*: [habja] (fricative despirantizes)

It is therefore not completely obvious that a (series of) diachronic changes has the same status as a phonological alternation. On the one hand, certain processes seem more common as diachronic processes than as synchronic phonological alternations (e.g. Compensatory Lengthening). On the other hand, Tswana shows that some alternations do not correspond to a single diachronic change.

3.5. Pathologies

- **Applicable to:** (So far mostly) static patterns
- **Example:** Aphasic patients (native speakers of Dutch) prefer to keep the obstruent when simplifying an obstruent + sonorant onset (Ouden 2002)
- **Pro:** May show unique evidence about the relation between phonology and brain;
- **Contra:** So far, there is little evidence that language areas in the brain are completely specialized, and they are always close to other areas, so that it is not always clear whether a pathology is really 'linguistic' or rather a problem of motor control.

Jakobson (1942) was one of the first to point out the potential relevance of pathological data (in his case, aphasia) for phonological theory. His point was that phonological markedness was reflected in aphasic speech — the less marked structures would be more resistant to the damaging effects of aphasia than the more marked ones. The topic has received some modest attention in the literature (see Buchwald 2008:for a recent overview).

Another type of pathology which is potentially relevant is dyslexia. This topic has received more attention, but mostly from a psychologist's perspective. Linguists however do pay some attention to the issues involved as well (for instance de Bree 2007).

A third type is what we could call 'everyday pathologies': speech errors (Fromkin 1973, Stemberger and Stoel-Gammon 1991, Goldrick and Daland 2009). It is often observed that the results of e.g. metatheses and substitutions in speech errors still yield well-formed words in the language of the speaker. To the extent that such is true, these errors can be seen as evidence for the working of the grammar — for the 'productivity' of phonology.

Data of this type are then very interesting — also because they may yield a unique window on the working of language in the brain, at least as long as brain scanning techniques are still outside the reach for a lot of theoretically inspired linguistic work — but still not very popular. One reason for this may be that

they are usually messy and very difficult to judge, among other things because a stroke hardly ever affects only one region in the brain; similarly, the origin of dyslexia apparently can be manifold.

3.6. Language acquisition

- **Applicable to:** Anything
- **Example:** Dutch children at some stage say *toto* instead of *oto*
- **Pro:** Since first language acquisition is a central question in most cognitively inspired views of language, the way it functions can shed light on the organisation of language
- **Contra:** Data is chaotic, evidence often very indirect.

The phonological study of language acquisition is much more developed than that of pathologies, but even here there are many things not very well understood. Typically, one distinguishes between first and second language acquisition (with possibly third language acquisition as an interesting extra category).

First language acquisition seems the most directly relevant to phonological theory. Obviously, since the advent of generative grammar the way in which a language is acquired has been put forward as evidence for aspects of phonological theory — indeed, the possibility of acquisition itself under an impoverished input is a traditional argument in favour of the generative enterprise.

The study of second language acquisition is a rather lively field in its own right.

4. Creating Data 1: Artificial Evidence

The types of evidence discussed in the previous section were 'natural' in the sense that they dealt with types of data which basically already exist before the researcher starts noticing them. A lot of data however is the result of actual intervention of the researcher: data which would not exist without the study which reports on them.

Creating such data is the bread and butter of almost every discipline, except maybe the historical sciences, since it allows a discipline to make real *predictions*, not just about data which are already known.

I divided the types of 'created data' into two subgroups. In this section, I first discuss what I call 'artificial evidence', which means data that are like those in the previous sections, except that the stimuli have been made up by the investigator. In the next section (5) I discuss the type of evidence that comes from experiments as we know them from the (social) sciences.

4.1. nonse words

- **Applicable to:** Anything phonological
- **Example:** Where do English speakers put stress in *totarentanton*?
- **Pro:** A test for the productivity of phonological processes for which not enough morphological material is available; no interference from diachrony possible by definition

- **Contra:** Possible interference from other psychological processes (the nonse words might be treated as foreign, etc.)

nonse words are traditionally an important test for linguistics at least since Gleason (1958) introduced the *wug* test: a child would get a nonsense noun such as *wug* and was asked to give the plural of this form. It was then checked whether the 'right' allomorph was chosen (in this case, a voiced [z]). *Wug* tests belong since then to the standard repertoire of the analyst of the phonology - morphology interface. Without it (when also applied to adult subjects), it would be difficult to find the relevant data for productive phonological processes, given that it is not completely unlikely that, in this case, speakers have stored also regular plurals in their memory.

Similar tests can obviously also be performed on phonotactics or stress. As evidence for the fact that speakers have intuitions about these things, Halle (1962) mentions the triple *brick, blick, bnick*, of which *brick* is an actual word of English, whereas there is a contrast between *blick* and *bnick*: although neither of these words exist, speakers have the intuition that *blick* is at least a possible word, whereas *bnick* is not.

Bakovic (2007) observes that *blick* actually does occur in English dictionaries, although it is unknown to most native speakers — which is the relevant criterion. But Bakovic (2007) makes another observation as well: " both phonotactically possible but allegedly nonexistent *blick* and phonotactically impossible and supposedly nonexistent *bnick* have by sheer force of repeated use become words of my vocabulary". The problem is that an acceptability judgement is subject to many complicated factors of which access to the grammar is one, but probably not the only one (Carstairs-McCarthy 1999, Kawahara 2011).

4.2. Artificial learning

- **Applicable to:** Any claim that is typological
- **Example:** Can human beings learn a system that displays across-the-board consonantal place harmony without vowel place harmony?
- **Pro:** A way to answer the question whether many / most / all typological gaps might be due to historical accident
- **Contra:** The tests are usually done with adults without the normal motivation to learn a language; it is difficult to exclude that other cognitive faculties are used.

A type of evidence that has become more popular recently is the artificial learning of phonological patterns. These are for language typology what nonse words are for the phonology of individual languages: a way to distinguish accidental gaps from 'real' gaps, caused by the properties of human phonology. If a theory predicts that certain patterns should be possible in human language, and we do not find any actual language that displays the pattern, we can try to teach the pattern to a group of subjects and compare their results to that of a control group which learns a pattern which is (ideally) formally very similar but would be predicted to be impossible in natural language. For instance, we can teach one group of subjects words that obey the Latin stress rule, and another group a set of words with stress on the final prime numbered syllable.

Moreton and Pater (2011) provide a solid overview of the use of artificial language in phonology, giving several examples and discuss the issues which arise (see also Carpenter 2008). Primary among this is the question whether subjects apply 'domain-general' mechanisms rather than phonology to the question

at hand. Many patterns can be learned by subjects applying certain general cognitive processes, which we then have to distinguish from 'real' language acquisition. This can be done in several ways, e.g. by showing that the 'non-natural' language is learned by a different group of people (there is a different correlation with general intelligence), or people make different kinds of mistakes, or they use different parts of their brains.

One problem is that these 'languages' are still quite different from 'real' human languages in a number of ways. They have no community of speakers, there is no way to use them in everyday situations and they usually also do not have a semantics or even a syntax if they are designed for phonological experimentation: subjects are simply subjected to sequences of meaningless sounds, and at least according to some definitions, those are not linguistic objects. Nevertheless, we would have to explain consistent differences between the 'natural' and 'unnatural' patterns if they both have been offered in a non-linguistic context.

4.3. Language games and secret languages

- **Applicable to:** Mostly claims about prosodic structure
- **Example:** Goat Latin
- **Pro:** Same as for nonse words, except that these games have arisen 'spontaneously' and have been tested by real language users.
- **Contra:** Language games often use processes that are completely weird from a typological perspective (e.g. metathesizing syllables).

Language games and secret languages have been the object of some study in the 1990s, usually in the context of Prosodic Morphology (McCarthy 1991, Bagemihl 1988; 1995, Vogt 2010). The words in language games and secret languages can be seen as nonse words that have arisen more or less spontaneously – or at least without the intervention of a linguist or psychologist.

Furthermore, one can make a typology of language games, and test phonological tools on those. (McCarthy 1991), for instance, shows that in languages with 'real' long vowels, those vowels can be split up by infixes. Slovak has an infixing secret language, which puts *pV* after every vowel, with *V* a copy of that vowel. Long vowels are not treated as units (Birnbaum 1981):

(3) a. nemaːm – nepemapam
b. netʃiːtam – nepetʃipitapam
c. choraː – choporapa

This can be taken as evidence for a prosodic representation of long vowels as one unit linked to two positions. English secret languages, such as Goat Latin (in which one puts *bəwV* inside of every syllable, do not show the same behaviour, which may be taken to indicate that the English diphthongs are not 'really' long vowels:

(4) ðɪs ɪz gowt lætɪn – ðɪbəwɪs ɪbəwɪz gowbəwowt læbəwætɪbəwɪn, *gobəwowt

The fact that there are such systematic differences between languages is of course interesting and should be indicative of the knowledge of the speakers. An obvious problem in interpreting language games and secret languages is that they are still consciously constructed, even if not by linguists, and that they often involve practices which are not found in 'natural' language. For instance, a common procedure is to reverse the order of syllables or segments in a word — a kind of extreme metathesis that no language employs in

normal morphology. Given that we allow for such 'unnatural' operations in the definition of the game, the issue is to what extent we can be sure that the responses to those are necessarily natural.

4.4. Poetry

- **Applicable to:** Claims about prosodic structure (maybe segmental things, in the study of rime).
- **Example:** Conclusions about syllable weight in Ancient Greek and Latin.
- **Pro:** At least some poetic traditions are arguably spontaneous or at least not initiated by literate people; they seem to reflect real patterns and are sometimes our only window into prosody
- **Contra:** Human agents can do anything they like in artistic production, there are no clear limits on what would be 'too artificial' to warrant serious linguistic study.

The study of poetry, and in particular of metrics, has gone hand in hand with that of phonology in general. For written languages which have died out before they were ever recorded, the systematic sound patterns in poetry are often among the few indications we have of the phonological structure of those languages. For instance, almost everything we know about Greek accent, we know from the study of poetry (Steriade 1982; 1988, Wetzels 1986, Golston 1990, Kiparsky 2003). Similarly, we can detect whether phonemes merged in a certain period by studying rime patterns. Labov (1972) uses folk rhyme to study patterns of variation in vowel quality.

But the poetic structure of living languages is also studied by phonologists. Very similar things can be said about this type of study as we said about language games in section 4.3: on the one hand these systems are artificial in the sense that they are the result of conscious intervention and linguistics tends to study mental mechanisms that are subconscious. In particular modernist poetry, in which poets consciously tried to break all rules, have received very little linguistic attention.

At the same time, we can assume that certain artificial patterns will be more successful than others (poetic metrics has binary and ternary but no quadernary feet) and relate this to subconscious linguistic knowledge.

5. Creating Data 2: Experimental Evidence

According to Coleman (2011), the only reliable evidence we have in any scientific discipline is experimental evidence, acquired under heavily controlled conditions in a laboratory. Under such a point of view, data acquired from introspection, fieldwork or corpus study are all unacceptable. While for many phonologists, this statement is too strong, there is definitely something to be said in favour of laboratory conditions: we can filter out interferences from many factors which do not interest us. At the same time, the artificiality of such conditions is also sometimes problematic: speakers will speak less naturally if they are put behind a microphone. This is Labov's well-known Observer's Paradox (Labov 1972; 1977; 2006)

5.1. Phonetic measurement

- **Applicable to:** Static phenomena, mostly segmental (or suprasegmental)
- **Example:** Final devoicing in German is not absolutely neutralizing
- **Pro:** We find patterns that may not be directly consciously accessible, but still have a bearing on the theories; the phonology presumably must be somewhere in the signal.
- **Contra:** at the same time, the phonetic signal is 'polluted' by all kinds of other factors (sociolinguistic, etc.); we do not know *where* the phonology is in the signal.

Most introductions to phonology (as well as to phonetics) start with making an explicit distinction between phonetics and phonology. There is a rather long list of criteria the distinguish the two, although not all of these converge: phonetics is about the physiological and physical aspects of speech sound whereas phonology is about psychological or sociological reality; phonetics is about continuous measures, whereas phonology is discrete. So what do we do with phenomena which seem clearly located in the human mind yet display aspects of continuity?

No matter how this may be, phonetics has developed a rich set of measuring techniques, which can be used to test phonological theories, if applied with care. An example is the demonstration by a number of authors that final devoicing is non-neutralizing in a number of languages (see Port and Leary 2005:and references cited there). It is virtually impossible for speakers to become aware of the trace of voicing which is left in a devoiced obstruent, but phonetic measurement can show that it is there — in a laboratory condition, which in itself has been the topic of a lot of argument. Furthermore, listeners seem to be able to pick up on those cues, albeit not perfectly.

Let us assume for the sake of the argument that these results are true. This must mean something for phonological theory; at least that a simple model in which indeed the difference between a voiced and a voiceless segment is completely neutralized and that the interface with phonetics treats them as one and the same object is false. This would then be an aspect of phonology which is simply not directly available for introspection, either by a linguist or a speaker.

Possibly the biggest problem in interpreting phonetic data is that the speech signal contains a lot of non-linguistic information — about the quality of the voice, about the age, gender and sociolinguistic characteristics of the speaker, about her emotional state, etc. (Kaye 1989). It is often very difficult to separate these factors and to know which parts linguistic theory proper is supposed to account for. Inversely, it is not always clear where to look for phonetic reflexes for abstract categories posited by phonological theory; but this obviously does not mean that such categories are not present: we may just not know where to look for them.

5.2. Psycholinguistic experimenting

- **Applicable to:** Mostly dynamic phenomena, but also patterns
- **Example:** French listeners find it easier to recognize *pa* than *pal* in *palace*, but the other way round in *palmier*. (Cutler et al. 1986)

- **Pro:** Again, we can find patterns that are not directly consciously accessible.
- **Contra:** Again, many unknown factors are involved in psycholinguistic experimenting; furthermore, unfortunately there is not a lot of work done that seems theoretically informed.

A different type of experimentation might be called psycholinguistic — although the boundaries with phonetic experimentation are not very clear. I will keep the discussion about this rather short, as it has been the topic of a lot of dicussion, for instance in Lahiri and Reetz (2010), Cohn et al. (2012) For the same reason, I will not go into other types of experimental methods one might further distinguish, such as brain scanning, as these data do not seem to pose different types of problems from phonetic or psycholinguistic experiments and offer similar advantages of providing objective evidence on phenomena that are not open to introspection. See e.g. Eulitz (2007), Idsardi (2007).

An important difference between phonetics and psycholinguistics is that obviously the latter is concerned more directly with mental patterns and in that sense it might be expected that psycholinguistic data would usually be more directly relevant for phonology. At the same time, also psycholinguistic patterns probably are influenced by a lot of non-phonological factors (having to do with the setup of the experiments, for instance), which are not always well-understood. One might say that a further problem is that a lot of psycholinguistic work is done without much understanding of phonological theory, but this is obviously not an argument against the methodology itself, which seems to me in principle one with a lot of potential.

6. Beyond Data: Formal evidence

A final type of evidence is one which does not consist of language data, but instead of data about the structure of the theory itself. This is the point where we move from Prince (2007)'s Theory of Data to his Free-Standing Theory. In most phonological practices, this type of evidence bears weight especially in the comparison of different theories.

6.1. Formal simplicity

- **Applicable to:** Overall grammars, individual analyses.
- **Example:** Assuming that final devoicing in German is lexicalized, would mean that thousands of nouns and verbs have to be stored in more than one allomorph.
- **Pro:** This is a classical cornerstone of epistemology and the philosophy of science.
- **Contra:** It seems impossible to create a theory-dependent way of counting simplicity.

Formal simplicity is the main, and probably the only, criterion for distinguishing two theories which cover the same empirical ground. This type of evidence is furthermore very often invoked in the literature, arguing that some proposal gives a 'more elegant' account of a phenomenon than some other proposal. Since elegance at least in mathematical proof seems to be equivalent to formal simplicity, we may assume that the phonological claims should also be subjected to a requirement of simplicity; but such arguments are hardly ever substantiated.

Maximal simplicity is a requirement for any serious theory of anything since the days of William of Ockham. In phonology, it seems to have been worked out only in Chomsky and Halle (1968) in the form of an 'evaluation metric', which stated that the best grammar is the one which uses the smallest number of elements. As Prince (2007) points out, this exercise seems never to have been performed over larger fragments of grammars, only very locally (in the comparison of one or two rules).

Chomsky and Halle (1968)'s metric was intended to compare two grammars written within the same formalism. Things become even more difficult, if not outright impossible, if one tries to to objectively compare two completely different formalisms. Comparison of this type therefore in the end is a matter of taste. Although it is possible to sometimes reach intersubjective agreement, and obviously it is always a good idea to stay as close as possible to formal simplicity, it is risky to use the relative complexity of a theory as an argument against (or in favour of) it.

6.2. Modeling

- **Applicable to:** Anything (in particular dynamic aspects of phonology).
- **Example:** Nevins (2010) algorithm for deriving vowel harmony.
- **Pro:** Requires maximal explicitness of the theory; potentially allows for verification of very complex predictions.
- **Contra:** Not every theory can be easily modeled.

A final type of evidence we can mention here comes from modeling: it proves the value of a theory if we can implement it in a working computer model. There are two reasons why this counts as evidence in favour of a theory. One is that a computer model obviously requires maximal explicitness: nothing can be left to the reader to figure it out, all assumptions have to be built into the computer program in one way or the other. The other reason is that theories can often be very difficult to verify by pencil and paper: many factors start interacting and it becomes almost impossible for the human mind to oversee them.

Nevins (2010) gives an example of this: accompanying his book-length study on vowel harmony is a computer script which implements the algorithm assigning stress. The algorithm can be tested and the computer code underlying it can be inspected (and compared to the pseudocode in the book). The advantage is that in this way one can, given enough effort, completely understand the theory and its predictions.

Obviously, having a computer model is not a requirement for a good theory, nor is it even an argument in favour against one. Furthermore, certain theories might just be more difficult to implement as a computer model than others (for instance ones which are based on parallel computation might in practice be very inefficient on present-day computers) and since we do not know that the human brain works like a digital computer there is no demand that theories are implementable. However, if a working model exists for a given theory, this is definitely an argument in its favour.

7. Conclusions

What counts as relevant evidence for a theory of phonology, if we define that field as the study of the knowledge of sound patterns in human language? It seems to me that there is no way to a priori exclude the relevance of any kind of evidence, mentioned here, for the study of phonology. There might obviously also be other types of evidence, yet to be discovered, which can throw light on the topic of our study.

Every kind of evidence should at the same time be taken critically, none is without conceptual problems and furthermore there simply is *no* direct window at present to the human mind. Every piece of evidence we have is therefore potentially polluted by other information. It is therefore very important to always realize what a piece of evidence is supposed to argue for, and how the data could be explained in an extra-phonological way. The best type of evidence is converging evidence: if several types of evidence mentioned above converge in showing the correctness of a given theory.

However, an important conclusion to draw is that we are lucky that so many types of evidence become available: a discipline which allows itself to build on such a large set of resources will thrive.

I believe this is good news in particular also for theoretical phonology. The different types of evidence mentioned here belong to very different ontological domains. In itself the result of a phonetic measurement has nothing to do with a frequency count in a corpus — these things belong to completely different realms of reality. We can only compare them if we move to a level that is more abstract than each of them. This is sometimes overseen by authors arguing for more 'concrete' forms of phonological representation: that there are many different types of concreteness, and there seems no inherent reason to prefer one over the other.

References

Bach, Emmon and Robert T. Harms. 1972. How do languages get crazy rules? In *Linguistic change and generative theory. Essays from the UCLA Conference on Historical Linguistics in the Perspective of Transformational Theory, February 1969*, edited by Robert P. Stockwell and Ronald K.S. Macaulay, pp. 1–21. Indiana University Press, Bloomington.

Bagemihl, Bruce. 1988. The morphology and phonology of Katajjait (Inuit throat games). *Canadian Journal of Linguistics* 33: 1–58.

Bagemihl, Bruce. 1995. Language games and related areas. In *The Handbook of Phonological Theory*, edited by John A. Goldsmith, pp. 697–712. Blackwell, Cambridge, Mass., and Oxford, UK.

Bakovic, Eric. 2007. Did you mean: brick block brick.

Bakovic, Eric. 2011. Opacity deconstructed. To appear in the second edition of The Handbook of Phonological Theory (Goldsmith et al., ed.).

Barbiers, Sjef. 2009. Locus and limits of syntactic microvariation. *Lingua* 119 11: 1607–1623.

Bateman, Nicoleta. 2007. *A crosslinguistic investigation of palatalization*. Ph.D. thesis, University of California, San Diego.

Becker, Michael, Andrew Nevins, and Nihan Ketrez. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in turkish laryngeal alternations. *Language* pp. 84–125.

Beckman, Mary E., Kiyono Yoneyama, and Jan Edwards. 2003. Language-specific and language-universal aspects of lingual obstruent productios in japanese-acquiring children. *Journal of the Phonetic Society of Japan* 7: 18–28.

Birnbaum, D. 1981. Rising diphthongs and the slovak rhythm rule. In *Harvard Studies in Phonology II*, edited by G.N. Clements, pp. 1–16. Indiana University Linguistics Club, Bloomington, IN.

Boersma, Paul. 2004. A stochastic ot account of paralinguistic tasks such as grammaticality and prototypicality judgments. ROA 648.

de Bree, Elise Henriëtte. 2007. *Dyslexia and Phonology. A study of the phonological abilities of Dutch children-at-risk of dyslexia*. Ph.D. thesis, Universiteit Utrecht.

Buchwald, Adam. 2008. Minimizing and optimizing structure in phonology. evidence from aphasia. *Lingua* 119: 1380–1395.

Burzio, Luigi. 2011. Derived environment effects. In *The Blackwell Companion to Phonology*, edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, vol. 3, chap. 88, pp. 2089–2114. Wiley-Blackwell, Oxford, UK.

Carpenter, Angela. 2008. *Learning Artificial Languages: The Role of Universal Grammar*. VDM Verlag Dr. Mueller e.K.

Carstairs-McCarthy, Andrew. 1999. *The origins of complex language: an inquiry into the evolutionary beginnings of sentences, syllables and truth*. Oxford University Press, Oxford.

Chomsky, Noam. 1957. Review of fundamentals of language. *International Journal of American Linguistics* 23: 234–242.

Chomsky, Noam. 1964. *Current issues in linguistic theory*. Mouton, The Hague.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. MIT Press, Cambridge, Mass.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English*. Harper & Row, New York.

Cohn, Abigail C., Cécile Fougeron, and Marie K. Huffman (eds.) . 2012. *The Oxford Handbook of Laboratory Phonology*. Oxford Handbooks in Linguistics. Oxford University Press.

Coleman, John. 2011. A history maker. *Journal of Linguistics* 47: 201–217.

Cutler, Anne, Jacques Mehler, D. Norris, and J. Seguí. 1986. The syllable's differing role in the segmentation of french and english. *Journal of Memory and Language* 25: 385–400.

Durand, Jacques, Ulrike Gut, and Gjert Kristoffersen (eds.) . 2012. *Handbook of Corpus Phonology*. Oxford University Press, Oxford.

Eulitz, Carsten. 2007. Representation of phonological features in the brain: evidence from mismatch negativity. In *Proceedings of 16th International Congress of Phonetic Sciences*.

Fromkin, Victoria (ed.) . 1973. *Speech errors as linguistic evidence*. Mouton.

Gleason, Jean Berko. 1958. The child's learning of english morphology. *Word* 14: 150–77.

Goldrick, Matthew and Robert Daland. 2009. Linking speech errors and phonological grammars: insights from Harmonic Grammar networks. *Phonology* 26: 145–185.

Golston, Chris. 1990. Floating H (and L*) Tones in Ancient Greek. In *Proceedings of the Arizona Phonology Conference*, vol. 3.
Hale, Mark and Charles Reiss. 2008. *The Phonological Enterprise*. Oxford University Press, Oxford.
Halle, Morris. 1962. Phonology in generative grammar. *Word* 18: 54–72.
Hay, Jennifer and Katie Drager. 2010. Stuffed toys and speech perception. *Linguistics* 48 4: 865–892.
Hume, Elizabeth. To appear. *Phonological markedness*. Key Topics in Phonology. Cambridge University Press.
Hyman, Larry. 2007. Elicitation as experimental phonology: Thlantlang lai tonology. In *Experimental Approaches to Phonology*, edited by M.J. Solé, P. S. Beddor, and M. Ohala. Oxford University Press, Oxford, UK.
Hyman, Larry M. 1975. *Phonology. Theory and analysis*. Holt, Rinehart and Winston.
Idsardi, William J. 2006. Canadian raising, opacity, and rephonemicization. *The Canadian Journal of Linguistics* 51 2/3: 119–126.
Idsardi, William J. 2007. Some meg correlates for distinctive features. In *Proceedings of 16th International Congress of Phonetic Sciences*.
Inkelas, Sharon, Cemil Orhan Orgun, and Cheryl Zoll. 1996. Implications of lexical exceptions for the nature of grammar. In *Derivations and Constraints in Phonology*, pp. 393–418. Clarendon Press, Oxford, UK.
Jakobson, Roman. 1942. *Kindersprache, Aphasie und allgemeine Lautgesetze*. Lundequist, Uppsala.
Joos, Martin. 1942. A phonological dilemma in canadian english. *Language* 18: 141–144.
Kavitskaya, Darya. 2002. *Compensatory lengthening. Phonetics, phonology, diachrony*. Routledge, New York and London.
Kawahara, Shigeto. 2011. Modes of grammaticality judgement. Ms, Rutgers University.
Kaye, Jonathan. 1989. *Phonology. A cognitive view*. Cognitive Science Series. Tutorial Essays. Lawrence Erlbaum Asociates, Hillsdale, New Jersey and Hove and London.
Kaye, Jonathan D. 1995. Derivations and interfaces. In *Frontiers of Phonology*, edited by Francis Katamba and Jacques Durand, pp. 289–332. Longmans.
Keller, Frank and Ash Asudeh. 2002. Probabilistic learning algorithms and optimality theory. *Linguistic Inquiry* 33: 225–244.
Kiparsky, Paul. 2003. Accent, syllable structure, and morphology in Ancient Greek. In *Selected papers from the 15th Symposium on Theoretical and Applied Linguistics*, edited by Elizabeth Mela Athanasopoulou. Thessaloniki.
Kochetov, Alexei. 2011. Palatalisation. In *The Blackwell Companion to Phonology*, edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, pp. 1666–1690. Wiley-Blackwell.
Labov, William. 1972. *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia.
Labov, William. 1977. *What is a linguistic fact?*. Peter de Ridder Press.
Labov, William. 1987. Some observations on the foundation of linguistics. Available at http://www.ling.upenn.edu/~wlabov/Papers/Foundations.html.
Labov, William. 1996. When intuitions fail. In *Chicago Linguistic Society. Papers from the Parasession on Theory and Data in Linguistics*, vol. 32, pp. 77–106. CLS.
Labov, William. 2006. *The Social Stratification of English in New York City*. Cambridge University Press, Cambridge, 2nd edn.
de Lacy, Paul. 2009. Phonological evidence. In *Phonological argumentation: Essays on evidence and motivation*, edited by Steve Parker. Equinox.
Lahiri, Aditi and Henning Reetz. 2010. Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics* 38: 44–59.
Lightner, Theodore. 1972. Some remarks on exceptions and on coexistent systems in phonology. In *The Slavic Word*, edited by D. S. Worth, pp. 426–442. Mouton, The Hague.
McCarthy, John J. 1991. L'infixation reduplicative dans les langages secrets. In *Les javanais*, edited by M. Plenat, pp. 11–29. Larousse, Paris. Langages 101.
Moreton, Eliott and Joe Pater. 2011. Learning artificial phonology. a review. Ms, UNC Chapel Hill and University of Massachusetts at Amherst.
Nevins, Andrew. 2010. *Locality in Vowel Harmony*. No. 55 in Linguistic Inquiry Monographs. The MIT Press, Cambridge, MA.
Ohala, John. 1986. Consumer's guide to evidence in phonology. *Phonology* 3: 3–26.
Ouden, Dirk Bart Den. 2002. *Phonology in Aphasia: Syllables and segments in level-specific deficits*. Ph.D. thesis, University of Groningen.
Port, Robert and Adam Leary. 2005. Against formal phonology. *Language* .
Prince, Alan. 2007. The pursuit of theory. In *Cambridge Handbook of Phonology*, edited by Paul de Lacy. Cambridge University Press.
Schütze, Carson. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. University of Chicago Press, Chicago.
Simon, H. and H. Wiese (eds.) . 2011. *Expecting the Unexpected*. Mouton de Gruyter.
Smith, Jennifer. 2001. Lexical category and phonological contrast. In *Papers in Experimental and Theoretical Linguistics 6: Workshop on the Lexicon in Phonetics and Phonology*, edited by Robert Kirchner, Joe Pater, and Wolf Wikely, pp. 61–72. University of Alberta, Edmonton.
Smith, Jennifer. 2011. Catgory-specific effects. In *The Blackwell Companion to Phonology*, edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, vol. IV, pp. 2439–2463. Blackwell-Wiley, London.
Stemberger, Joseph and Carol Stoel-Gammon. 1991. The underspecification of coronals: evidence from language acquisition and performance errors. pp. 181–200. (1991).
Steriade, Donca. 1982. *Greek prosodies and the nature of syllabification*. Ph.D. thesis, MIT, Cambridge, Mass.
Steriade, Donca. 1988. Greek accent. a case for preserving structure. *Linguistic Inquiry* 19: 271–314.
Vogt, Barbara Maria. 2010. *Prosodische Morphologie in Geheimsprachen. Phonologische Manipulationen als Affigierungs- und*

Reduplikationsprozesse. Ph.D. thesis, University of Verona.
Wetzels, Leo. 1986. Phonological timing in ancient greek. In *Studies in Compensatory Lengthening*, edited by L. Wetzels and E. Sezer, pp. 279–344. Foris, Dordrecht.
Wimsatt, William K. and Monroe C. Beardsley. 1946. The intentional fallacy. *Sewanee Review* 54: 466–488.
Yoneyama, Kiyono, Mary E. Beckman, and Jan Edwards. 2003. Phoneme frequencies and acquisition of lingual stops in japanese. Ohio State Universiry.
Zonneveld, Wim. 1978. *A Formal Theory of Exceptions in Generative Phonology*. Foris, Dordrecht.
Zwicky, Arnold M. 1980. 'internal' and 'external' evidence in linguistics. *PSA* 2: 598–604.