Evaluating Evidence for Stress Systems*

Paul de Lacy
Rutgers, The State University of New Jersey
Version: October 1, 2012

delacy@rutgers.edu
http://www.pauldelacy.net
Linguistics Department
18 Seminary Place
New Brunswick, NJ
08901
USA

To appear in Harry van der Hulst (ed.) (to appear). Word stress and typological issues (working title). Cambridge University Press.

1 Introduction

The core of Echeverría and Contreras's (1965) (EC) description of Araucanian/Mapudungun[1] stress is given in (1).

(1) *Echeverría and Contreras (1965: 134)'s Araucanian/Mapudungun Stress Description*
"a. GENERAL RULE: a phonological word has a main stress on the second syllable and, if applicable, secondary stresses on the fourth and sixth syllables: /wulé/ *tomorrow*, /ṭipánto/ *year*, /elúmuyù/ *give us*, /elúaènew/ *he will give me*, /kimúfalùwulày/ *he pretended not to know*."

EC's description has been cited as evidence for a variety of metrical theories (e.g. Hyman 1977: 41-42,75, Kager 1993: 409, Hung 1994: 177-180, Hayes 1995: 266, Gordon 2002: 522, Hyde 2002: 320-325, McGarrity 2003: 59-61, Tesar 2004:220-221). As a source of stress evidence, it is fairly typical. Large stress studies such as Hayes (1995) and Gordon (1999) overwhelmingly rely on descriptions like EC's (i.e. journal articles, grammars). In Gordon's (1999: appendix 1) list of 392 stress systems the mean publication date for earliest sources is 1972 and the median is 1975, so EC is slightly older than the average source.

The issue addressed in this chapter is whether EC's description *should* be used as evidence for phonological theories.

The general issue – how to evaluate phonological evidence – has been addressed implicitly and explicitly before (e.g. Ohala 1986). A theme of this chapter is that the source of requirements on evidence is theories themselves. Different theories of phonology impose

* My thanks to Matt Gordon, José Ignacio Hualde, Harry van der Hulst, Larry Hyman, Luca Iacoponi, Catherine Kitto, and Benjamin Molineaux for their comments on this chapter, and to the audiences at Princeton (2008), Cornell (2011), and the 2nd *UConn Workshop on Stress and accent* at the University of Connecticut in 2011 for their comments on presentations closely related to this article.

[1] The metrical stress literature cited here uniformly refers to the language as 'Araucanian'. However, Zúñiga (2000:4) comments that "this name is rejected by both scholars and the indigenous organizations", so 'Mapudungun' will be used here.

different requirements on evidence. Here, the focus is on Generative theories of phonology (Chomsky and Halle 1968 and later theories in their framework).

Generative theories have a phonological module (PhM) that creates output representations that are translated by later modules into instructions to articulators; articulatory movements cause air perturbations which are perceived by hearers as speech sounds. The theory of production and perception of speech identifies many places in the process where recovery of the PhM's output can be 'distorted'. So, for any data, all sources of potential distortion must be examined to correctly recover the PhM's output state.

The distorting aspects of modules and mechanisms discussed in this chapter are well known; nothing new is proposed. The aim is instead to apply that knowledge and use it to determine evidentiary requirements. The examination will not be comprehensive – there are too many factors to allow detailed examination in so few pages. However, enough of the major distortion sources will be covered to illustrate the general approach.

In short, the aim here is to identify some of the requirements that Generative theories impose on evidence for states of the PhM, and determine whether EC's description satisfies those requirements.

An overview of the theory is given in section 2, and is used as a template for evaluating EC's description in section 3. I emphasize that this chapter is *not* a critique of EC, nor does it attempt to determine the stress system of Mapudungun. The goal is to evaluate whether EC's description is relevant and adequate evidence for a Generative metrical theory – a point discussed further in section 4.

The concerns expressed in this chapter relate to work on methodology of descriptions and analyses of stress. In this book, relevant discussion can be found in the chapters by Hualde and Nadeu, Hyman, and Gordon.

2. Theoretical framework

The theoretical framework employed here is the Generative theory of phonology, as proposed in Chomsky and Halle (1968) and developed in subsequent Generative theories (e.g. Lexical Phonology and Morphology – Kiparsky 1982, Optimality Theory – Prince and Smolensky 2004). There are many Generative metrical theories and subtheories, including Hayes (1981), Prince (1983), Kager (1993), McCarthy and Prince (1993), Hayes (1995), and Hyde (2002). Aspects of the system that are particularly relevant will be summarized here; the purpose of this section is to provide a framework for discussing the adequacy of EC's description in section 3.

At the core of the Generative framework is the Phonological Module (PhM) – a cognitive component that transforms symbolic representations. The PhM's input consists of strings of phonological symbols sourced from the Lexicon; at least some aspects of morphological and syntactic structure are visible to the PhM's transformations (e.g. Ussishkin 2007, Truckenbrodt 2007) (Fig. 1 does not make any claim about exactly how the morphological and syntactic modules relate to the PhM; it merely acknowledges their influence).

Every theory of the PhM specifies possible and impossible states of the PhM. For example, SPE and its successors defined possible rules and orderings/interactions (Chomsky and Halle 1968, Kiparsky 1982) while work in Classical OT specified constraints and possible rankings (Prince and Smolensky 2004).

Of course, the best evidence that a particular output state of the PhM exists would consist of phonological outputs. Unfortunately, so far there is no way to directly detect phonological outputs. Instead, the available devices measure the output of post-PhM modules and mechanisms. The most commonly reported type of evidence for PhM states is 'perceived speech sounds' (PSS): i.e. a hearer's perception of a speaker's speech sounds. PSS are far from the speaker's PhM output, as shown in Figure 1.

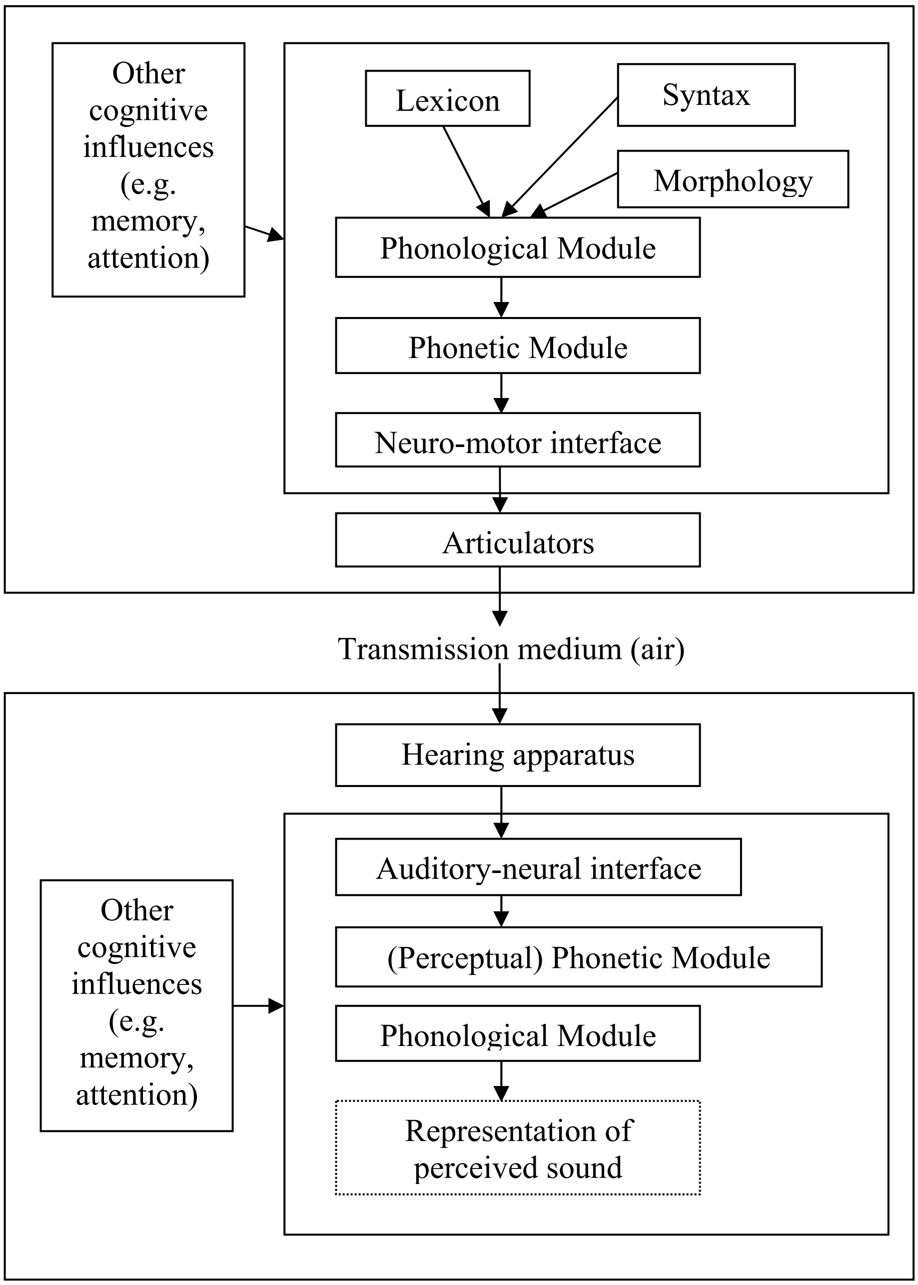

Fig 1. The PhM's role in generating human-perceived speech sounds

The PhM's output is sent to the Phonetic module, which creates a gradient representation that is implemented as articulator movements by the neuro-motor interface. The articulator movements create variations in air pressure and flow (e.g. Löfquist 1997). These perturbations reach the ear of the hearer, and his/her perceptual system forms a representation of what was perceived.

There are many modules and mechanisms between the speaker's phonological output and the hearer's perception. When using perceived speech sounds as evidence for PhM outputs, then, the contribution of all those modules and mechanisms to the perceived speech signal must be considered. It would be an easy task if all of the post-PhM modules and mechanisms were straightforward: i.e. they translated the output of the previous module/mechanism in a 1:1 fashion. However, it is likely that all of the modules/mechanisms are distorting in the sense that their translation can obscure input distinctions and introduce new elements to the speech signal, so making recovery of their input difficult.

The following section will expand upon the distorting effects of the modules/mechanisms and how such distortion affects the validity of evidence for the PhM. The focus is on perceived speech sounds because they are the most common type of evidence used for phonological theories: grammars and journal articles typically use descriptions that report an analyst's interpretation of their perception of a subject's speech sounds. Of course, perceived speech sounds are not the only source of evidence for the PhM. Idsardi (2010) provides an overview of neurophysiological techniques in measuring cognitive events. There are also 'intuitions' – a subject's verbal report of their beliefs about their speech (Chomsky 1965: 18ff; cf Kawahara 2011§1.2 and references cited therein).

It will be assumed that the analyst and subject are two different people. There may be a different process of perception in self-elicitation, but I am not aware of any studies that compare self-elicitation to elicitation from others. In some situations, the perceiver is a machine – a sound-recording and processing device. However, the same issues arise with mechanical recording devices as with humans: the machine's equivalent to an ear is a microphone, the auditory-neural interface is equivalent to a sound card, and the software used to transform the speech signal is analogous to the human's perceptual system (discussed in section 3.7), so similar distortion issues arise.

The final step is not represented in Fig.1 – where the hearer/analyst goes through a process of expressing the perceived speech sound in written form in an article or grammar. Aspects of the potential distortions produced at this final stage will be discussed in section 3.9.

3. Echeverría and Contreras (1965)

As mentioned in the introduction, EC's description of Mapudungun has been used as evidence for metrical theories in many publications. EC's description is significant: Hayes (1995: 71, 75) proposes that iambs are always quantity-sensitive, but EC apparently describes a quantity-*insensitive* left-to-right iambic metrical system (Gordon 2002: 522, 545). Section 3.9 below has further discussion. Apart from its theoretical significance, EC's description is so compellingly straightforward that it has often been used to exemplify left-to-right iambic rhythm (e.g. Kager 2007: 204-205).

The articles and books that cite EC's description do so for a theoretical purpose: they propose theories that predict a particular state of the PhM should exist, and EC's description is presented as evidence for that state. However, as there are so many influences on perceived speech sounds other than the PhM, EC's description must be examined to see if all those other distortions were taken into account.

The theme in the following sections is that the theory of the PhM and all the modules and mechanisms that influence it predicts where distortion can arise to obscure the PhM output's structure. So, the theory imposes requirements on the evidence that is presented for it in the sense that a comprehensive analysis will examine the potential sources of distortion and correct for them to recover the phonological output.

The following discussion does not aim to be comprehensive, advocate any particular techniques, or determine the metrical system of Mapudungun.[2] Instead, the aim is to selectively illustrate the point that the theory is the source of requirements on evidence by determining EC's value as evidence for Generative metrical theories.

However, one publication that must be mentioned is Echeverría (1964). The connection between EC and Echeverría (1964) is not entirely clear because neither work cites the other. However, it is possible that EC relied on Echeverría (1964)'s recordings to a great extent. Regardless, the goal of this chapter is to consider specifically whether EC alone provides adequate evidence; after all, EC is always the source that is cited in theoretical phonological literature on stress while Echeverría (1964) is not mentioned (except by Molineaux 2012). In addition, the two works present different stress generalizations, and slightly different stress patterns in their data (see section 3.9). Echeverría (1964) will only be discussed when a comparison between the two works is particularly illuminating.

3.1 The Lexicon

The forms in (2) are the ones usually presented to illustrate Mapudungun stress. A natural interpretation is that they show the work of a PhM process – i.e. metrical construction of non-exhaustive left-to-right iambs.[3]

(2) *Basic stress forms (EC p.134)*
(a) /wulé/ *tomorrow*
(b) /ṭipánto/ *year*
(c) /elúmuyù/ *give us*
(d) /elúaènew/ *he will give me*
(e) /kimúfalùwulày/ *he pretended not to know*.

However, the Lexicon can – in a sense – distort evidence for PhM mechanisms. The underlying phonological strings of lexical entries may specify stress position (e.g. Alderete 2001 and references cited therein). If the PhM preserves lexically specified stress then the stress reported in (2) may not be evidence for a metrical PhM process, but rather simply be preservation of underlying stress. There are therefore two potential sources for stress location: the PhM's metrification processes and the Lexicon.

EC present the forms in (2) as exemplifying a predictable process: "Stress is predictable with reference to the phonological word… Stress is accounted for by one general rule and four

[2] There are many other descriptions of Mapudungun stress. Section 3.9 identifies and discusses major grammars; Fabre (1998) and Zúñiga (2006:68) provide extensive bibliographies about Mapudungun phonology. The most recent work I am aware of is Molineaux (2012).

[3] Here and throughout data is cited verbatim from EC. My understanding of the IPA equivalents for EC's non-standard symbols is: ṱ = [t̪], ṭ = [ʈ], ṭʳ = [ʈʂ], ṋ = [n̪], ḽ = [l̪], ł = [ʎ], i̭ = [j]; /ï/ is [ɯ] in stressed position and [ə] in unstressed position.

special rules…" (EC 134). For a Generative phonologist, a natural interpretation of EC's comments is that there is a productive PhM metrification process, and the data provided is representative of the output of that process. However, do EC present adequate evidence for a metrification process? In (2), only one form is given for each word length. For example, only one form is cited for two syllables: /wulé/. Could this form have lexical stress, and the default metrification process produce initial stress, at least in disyllables? In some iambic systems disyllabic words are reported to have consistent initial stress, so it is not impossible that /wulé/ preserves a lexical stress while the default position is the initial syllable (e.g. Southern Paiute – Sapir 1930: 39).

So how many disyllabic forms would one need to be assured that the PhM metrification process assigned stress to the second syllable? The best answer may be "As many forms as the learner needs to converge on a metrical system with default final stress in disyllabic words". Unfortunately, while learners know exactly what the PhM is capable of and how many forms are needed to converge on a particular system, analysts do not (yet). So, other approaches must be used.

One might point out that if /wulé/ is a lexicalized form, then there should at least be *some* disyllabic form in which stress falls on the initial syllable; this idea follows from the assumption that if a learner was only exposed to σσ́ words and never heard σ́σ words, the learner would always opt for a metrical system that assigned second-syllable stress.

EC do in fact cite disyllabic words with initial stress: e.g. /fílu/ *snake*, /čáo/ *father*, /rúko/ *breast*. So, as both σσ́ and σ́σ words are cited, there are two possibilities: either /wulé/ reflects a general metrical process and /fílu/, /čáo/ and /rúko/ have lexical stress, or the opposite is the case.

It is interesting that while the majority of metrical analyses of EC's description imply or assert that the default stress on disyllables is on the second syllable, EC state in a 'Special Rule' that: "Two-syllable words ending in a vowel may be stressed on either syllable, except particles (adverbs, pronouns, prepositions) which are always oxytone" (p.134). It is possible that particles do not reflect the general metrical pattern but have a class-specific stress pattern. So, the original issue remains: for disyllabic non-particles, what does the default metrification process generate?

EC's description also raises a third option. The statement that CV(C)CV words "may be stressed on either syllable" could be interpreted as describing free variation (EC is interpreted this way by Hyman 1977: 41-42). EC cite "/pu rúka/ ~ /pú ruká/ *in the house*" which does suggest free variation, but the stress difference might instead be conditioned by context that is not provided in the description. Other words are cited without suggestion that their stress position might vary freely (the exception is *red* which is cited as both /kélï/ and /kelḯ/ – EC 132).

To pursue the issue of data quantity further, one might expect that the default metrical pattern will be reflected in the largest number of forms: i.e. if a relatively larger number of σσ forms had final stress, then the metrification process by default assigns final stress. Table (3) summarizes the disyllabic forms cited by EC, arranging them by stress pattern.

(3) *EC: Stress position on disyllabic forms that are not particles*

Shape	Initial stress	Final stress	Total
CV.CV	15	4	19
CV.CVC	0	15	15
CVC.CV	3	2	5
CVC.CVC	0	3	3
ALL	18 (43%)	24 (57%)	42

Focusing on CVCV words, initial stress is default by a majority criterion. However, is there *enough* data here: would much more data reveal that the majority have final stress? While learners can resolve this issue, exactly how they do it is currently unclear: is there a certain ratio where (for example) if there are twice as many final-stressed forms as initial-stressed ones the metrical process is fixed as final stress? At the very least, given the variation in stress position observed in the data, I suspect that most phonologists would expect a great deal more data to resolve the issue.

The table in (3) underscores the difficulty of inferring a pattern from limited data. Of 15 CV.CVC forms, all have final stress, suggesting a pattern. However, if 15 initial stressed CV.CV forms are not convincing of CV.CV initial stress, then one wonders whether 15 forms should be convincing that CV.CVC forms have final stress; it is possible that it is an accident of the small sample size that all the cited CV.CVC forms have final stress. The same point applies to CVC.CV and CVC.CVC forms. If in fact CV.CVC forms do have final stress, then EC's data do not match their generalization, and in fact the stress system is quantity-sensitive with CVC syllables heavier than CV.

For CVC.CV forms, three have initial stress and two have final stress (There are three particles – not included in the table above – that have final stress). However, one of the initial stress forms is a Spanish loan, and one of the final stress forms is, too. That leaves three forms: two with initial stress [wéṇṭʳu] *man* and /láfṭa/ *small*, and one with final stress /kofké/ *bread*.

So, given the data provided for EC, it seems that no conclusion about the metrical system can be drawn with certainty about (at least vowel-final) disyllabic forms; it is possible that some have initial stress and some have final, or that there is free variation, or both. There is little solid evidence on the basis of the data cited to determine which stress position is the default.

The majority of data cited by EC consists of disyllabic forms (46 forms); there are 14 longer forms, summarized in (4). Longer forms have the appearance of being much more regular in stress location. For example, all six trisyllabic CVCVCV forms have medial stress. While two 4-syllable and two 5-syllable forms are cited with different stress patterns, it is possible that secondary stress was simply not transcribed for some forms. EC (p.134.c2) comment that the only CV.CV́.CV.CV.CV̀C form provided has a suffix *–la*, which shifts stress away from a preceding syllable (/ḽaɲ́ímïwulày/ *he did not kill himself*; /y/ is possibly another suffix; EC do not give morpheme divisions). However, Echeverría (1964:51)'s only five-syllable form /ṭɯlkéaliwén/ 'crust' has the same stress pattern and does not have the negative morpheme.

(4) *Trisyllabic and longer patterns cited by EC*

Syllables	Stress pattern	Forms	Total
3	CV.CV́.CV	6	
	CV.CV́.CV̀C	1	
	CV.CV́C.CV	2	
	CV.CVC.CV́C	1	10
4	CV.CV́.CV.CV	1	
	CV.CV́.CV.CV̀	1	2
5	CV.CV́.CV.CV.CV	1	
	CV.CV́.CV.CV̀.CV	1	
	CV.CV́.CV.CV.CV̀C	1	3
6	CV.CV́.CV.CV̀.CV.CV̀C	1	1

However, there is still some inconsistency in stress position in longer words: primary stress falls on the penult in tri-syllabic /θuŋúlàn/ *I do not speak* (EC 134), but on the final syllable in [umau̯ tún] *to sleep* (EC 133; [u̯] is the labio-velar glide [w]).

Finally, it is important to emphasize that most of the shapes and stress cited above are for non-surface forms – i.e. forms that are intermediate to the derivation; processes such as glide formation can cause stress rearrangement (see section 3.9 below).

Given that there is lexically specified (or freely varying) stress for disyllables and some longer forms, it seems reasonable to require confirmation that longer forms do not show variation. The problem is that 10, or 3, or 2, or 1 forms may not be enough to capture any variation patterns, or be convincing that there is no variation in longer forms.

A further concern is whether the assumptions behind the descriptive categories are correct. The terms CV and CVC lump together different sonority values even though it is known that some systems are sensitive to sonority distinctions (de Lacy 2007). In fact, an early description of Mapudungun stress – Augusta (1903:2-4) – claims that stress avoided the lowest sonority vowel [ə]: e.g. *fótəm*, rather than the expected final-stressed *fotə́m* (cf. *pəllí*, **pə́lli*). Similarly, the description assumes that all codas will be treated in the same way, even though some languages distinguish weight based on the sonority of the coda consonant.

A final assumption is that there are stress domains longer than three syllables in Mapudungun. Morphemes seem to be maximally trisyllabic – all longer words cited by EC are apparently morphologically complex. In fact, Salamanca & Quintrileo (2009) state that words have maximally three syllables in the Tirúa dialect of Mapudungun. Limits on word size or stress domain size are known in a variety of languages (Ketner 2006). In Echeverría (1964)'s list of words and phrases, the transcriptions of almost all words longer than three syllables are accompanied by an intonation symbol, perhaps indicating that they are phrases rather than single words. It would therefore be valuable to know whether words longer than three syllables constitute a single stress domain (i.e. prosodic word), or whether they always consist of several domains. If domains are maximally trisyllabic, then there can be no evidence for alternating stress.

In summary, a closer inspection of EC's data leads to uncertainty about the stress system. There may be lexical stress or free variation, or both in different environments. It is unclear what the default metrical pattern is; a 'majority criterion' is not helpful due to the small amount of data cited. Of course, apart from the data there remains EC's prose description of stress location. However, without a clear understanding of how their description relates to the data, any reliance on their assertions would merely be an appeal to authority – that is to say, accepting EC's description as correct depends on one's personal confidence in the describers, rather than on the basis of evidence.

While the Lexicon can present challenges in determining metrical systems, it can also provide opportunities. If words could be identified that were sure to have no underlying lexical stress, the theory predicts that they would undergo the productive metrification pattern. For example, the stress pattern of loanwords might be used (e.g. Hayes 1995: 143 for Fijian). EC cite /sánču/ *pig* and /kensá/ *maybe* as loanwords from Spanish; unfortunately, these forms have the same stress as the source language (*chancho* [tʃáɲ.tʃo], *quizá* [kisá]), offering the possibility that the source language's stress was preserved (for discussion of the complexities of loanword stress adaptation, see Broselow 2009). A wug test could be revealing, as long as analogical influence was controlled for (Berko 1958; see Kawahara 2011§2.1.1 for discussion re. phonological research); EC do not report the stress for any novel words.

Modules other than the Lexicon can also thwart the straightforward interpretation of stress data. For example, a PhM can have more than one productive metrification process, with different patterns for different classes of word (e.g. English nouns vs. verbs vs. adjectives – Burzio 1994: 43ff); morphological boundaries also influence stress domains (e.g. Nespor and Vogel 1986: ch.4). While EC's examples are not accompanied by a morpheme-by-morpheme decomposition, it seems that the majority of long forms are morphologically complex: e.g. /elúaen/ *you will give me*, /elúmuyù/ *give us*, /elúafimi/ *you will give him*, /elúaènew/ *he will give me*, [mawɨ́nài] *it will rain tomorrow*, /kimúfalùwulày/ *he pretended not to know*. With the paucity of examples longer than three syllables, the issue of influence of morpheme boundaries cannot be usefully examined. However, there is some suggestion that morpheme boundaries *might* influence stress placement: EC comment that for /-la/ *negative*, "the secondary stress is shifted from the fourth to the fifth syllable" (/l̪aŋɟ́mïwulày/, */l̪aŋɟ́mïwùlay/ *he did not kill himself*). Smeets (1989: 61) identifies several suffixes with fixed stress, suggesting that they either have lexical stress or that their morpheme boundaries influence stress placement in such a way to ensure that they receive stress. Without morpheme-by-morpheme glosses, it is difficult to determine whether EC's data reflects morphological influence. Similar issues arise for the Syntax module.

In short, the theory of the Lexicon requires that the issue of lexical stress be examined in any set of data. There are a variety of techniques for doing so, such as citing a great deal of systematic data and using tests for generativity. EC's description underlines the importance of addressing the issue: they describe variation in stress position, especially in disyllables. However, there is inadequate data to be sure of the default metrification pattern, or of the generalizations about other aspects of stress (e.g. where primary and secondary stress falls, where CVC attracts stress, the interaction of morpheme boundaries with stress).

I add that EC provide more stress transcriptions than I have seen in most articles and grammars; I have seen some that have no data at all, and many that list only a few forms. It is only because EC provide so much data that the difference between their generalizations and the data presented can be seen: without adequate data there is no way to evaluate the evidentiary basis for the describer's generalizations.

3.2 The PhM: L1, L2 and fluency

As a linguistic module, the PhM has properties such as L1/L2 status and fluency (occasionally called 'native speaker' status).

There are significant differences between L1 (or 'native') and L2 (or 'non-native') modules (Bley-Vroman 1988, Cook 2000; for the L1 vs. L2 acquisition of stress see Archibald 1995). The challenge in investigating an L2 module is that it is unclear whether its output only reflects the workings of that particular module, or whether other computational resources have been used (such as L1 modules, other L2 modules, or general cognitive processes). So, the stress output of an L2 module may not reflect a single coherent metrical system, but rather a complex mixture of metrical systems and lexical specifications. Similarly, intuitions reported about an L2 by a speaker might not draw on the same cognitive resources as L1 judgments (Davies and Kaplan 1998). In contrast, it seems likely that the output of an L1 module is influenced only by that module; so, the output of an L1 module provides a clearer window into the possible states of human PhMs.

PhMs can also be 'fluent' or 'non-fluent' (or 'fully developed' and 'not fully developed'). For example, a child can learn a language but have his/her learning cut short, and so not develop a large vocabulary or be confident in giving intuitions about acceptability. This is the case for a subject I worked with who grew up in Germany. He learned German until the age of 11, then left to live in an English-speaking country where he heard no German (even his parents avoided speaking German to him). Although his German PhM is L1 by the critical period criterion, his vocabulary is small and he is unable to express intuitions about sentence and word acceptability.

Strictly speaking, fluency is irrelevant to the overall goal: to determine possible states of the PhM. If every stage of language development employs a possible state of the PhM, then a less than fluent speaker (e.g. a young child) still has a possible PhM state, and determining the form of that state helps achieve the goal. However, there are practical problems when using data from less-than-fluent adult speakers: they may be more likely to code-switch in elicitation tasks, or draw on vocabulary and phonological patterns from their dominant adult language. The result is a set of data drawn from different PhMs.

So, to use EC's description as evidence for the claim that the PhM has a state that generates left-to-right iambs, the L1 status of the subjects' PhM must be determined. The worst case scenario would be if EC's speakers' PhMs were all L2, and their stress was influenced by other PhMs (e.g. their L1 PhM) in such a way that it did not reflect any possible individual PhM metrical system. In this case, EC's description would not reflect a possible state of the PhM, but only be a collection of disparate data.

EC report the following about their subjects: "Our study is based on material elicited (and recorded) from five informants, all of them from the Cautín province in Chile." (EC 132). The article implies that all the informants speak Mapudungun. No further information about the subjects is given.

EC does not include an assertion that the speakers were native speakers. Techniques for assuring L1 status can involve the subjects answering questions about their linguistic history, such as the language they grew up speaking, or other externally verifiable criteria (e.g. Labov 2006: 118-119).

I am not claiming here that Echeverría and Contreras were unaware of the issues surrounding L1 status or that they did not use techniques for determining L1 status. The observation is that EC does not report L1 status or their techniques, and so the implication for those who wish to use EC's description for Generative theories is uncertainty.

In fact, there is clear evidence that Echeverría was aware of the importance of L1 status. Echeverría (1964:19) notes "El primer paso para realizar nuestro estudio fue ubicar los hablantes nativos que proporcionaran el material lingüístico necesario." (The first step in our research was to locate native speakers to provide linguistic data."). It is possible that EC relied on Echeverría (1964)'s data – they both mention 'five subjects'. Unlike EC, Echeverría (1964:19-20) provides a great deal more information: for each subject their name, age, occupation, residence, and linguistic information are provided. For example, José Ailío Huaracán was born in Cautín province, lived in Concepcíon for more than 10 years, was 35 years old at the time of fieldwork, had primary school education, and had learned Mapudungun from childhood ("Aprendio el mapuche desde niño" – Echeverría 1964:19).

This chapter is about EC's description and whether it can be used as evidence for Generative theories, so I am reluctant to delve too far into Echeverría (1964)'s description. However, it is striking how EC and Echeverría (1964) differ: EC provide almost no information about their subjects while Echeverría (1964) provides several important details that are crucial for Generative theories. Echeverría (1964:19-20) notes that four of the five subjects learned the language while they were a child – an essential condition of L1 status. Fluency was also a concern: Echeverría (1964:20) notes that two subjects spoke the language at home, and one spoke it frequently. (For two subjects, there is no indication of how often they spoke it. For example, while Lorenzo Chicahual learned the language as a child, he had lived in Temuco for decades and there is no comment on whether he spoke the language regularly.)

The current expectation that only native speakers be used as evidence for PhMs was not always universal. For example, in Williams' (1917) dictionary of Māori, the preface reports that the data was taken primarily from secondary sources (e.g. dictionaries, word lists, texts) compiled by Europeans (non-native speakers). In the 'Methods' section of the Preface, Williams comments:

(5) *Methods for ascertaining wordforms in Māori (Williams 1917)*

> "There are various methods of ascertaining the meaning of a Maori word … The most natural procedure is to enquire from an intelligent Maori of the older generation, or preferably from several such. But this course, now unfortunately seldom available, is not always free from the risk of error. Few Maoris can resist the temptation to oblige an enquirer rather than admit ignorance; and occasionally a Maori may, in all honesty, have been habitually misusing a word in his own tongue. Moreover, local usage in respect of words is frequently very curious and perplexing."

So, Williams (1917) considered native speakers to be an important but imperfect source of information – they were prone to 'frequently curious' misuse of their own language. The preference was for secondary sources compiled by non-native speakers (e.g. Colenso's dictionary, "Mr Atkinson's papers", Sir George Grey's collection, the Rev. Father Becker of Hokianga's word lists, Mr S. Percy Smith's word lists, and so on), including Mr Elsdon Best who had "intimate acquaintance with the Maori", and "supplied a very large number of words, new meanings, and examples, collected at first hand."

Currently, Mapudungun is a minority language even among the Mapuche: of 700,000 Mapuche, 200,000 people speak the language fluently; the majority are bilingual (Zúñiga 2000: 4, UNPO 2008). EC estimate that there were about 100,000 speakers in 1965 in Chile. With just the information provided by EC, it is not possible to have confidence that the speakers were native speakers. If Echeverría (1964)'s information is taken into account, it is at least clear that all of the speakers were bilingual. However, for L1 status the details are still murky. One of the subjects may not have learned the language young enough, and a few may not have spoken the language regularly for many years.

Even though four of the speakers said that they spoke the language as a child, this criterion alone is not necessarily an adequate diagnostic tool for L1 status. This point is underscored by Echeverría (1964:20)'s comment. One of the subjects – María Catrileo – listened to the recordings of another subject – Ailío. Her opinion, combined with Ailío showing 'repeated ignorance of vocabulary and inconsistent pronunciation', led to Ailío being dismissed. ("En un estado avanzado del analisis del material, fue necesario desestimar al informante Ailío por presentar repetida ignorancia de léxico y pronunciación inconsistente."). It is possible that Ailío was not a native speaker of Mapudungun. If so, the fact that he self-reported that he had learned the language since he was a child was an inadequate diagnostic for determining native speaker status.

In general, it is far from easy to be sure that a person is a native speaker. In my own fieldwork on the Polynesian language Māori, I found that a number of people said they spoke Māori fluently or natively but actually knew only a few words; by saying they spoke Māori they were instead expressing a strong emotional connection to Māori culture. Some people reported speaking the language as a child, but close questioning revealed that it was not their primary language, and it was used in a classroom situation and not among their peers (e.g. never while playing). In contrast, one native speaker was very reluctant to say she spoke Māori; she was concerned that she did not have enough prestige to be a research subject.

In summary, without a clear description of the method for determining L1 and fluency status, there can be no reasonable assurance that subjects are appropriate sources of PhM evidence, as required by the theory.

Even if a PhM is L1, its output might not give insight into how PhMs usually work if it has been impaired (e.g. Stemberger and Bernhardt 2007; Cummings 2008: 265ff; Bernthal et al. 2009). It is somewhat hard to tease apart impairments of the PhM from Phonetic module impairments, or even neuro-motor impairments (Haynes and Pindzola 2011: ch.6). However, a number of disorders have a phonological character in that they refer to phonological restrictions (e.g. den Ouden 2002). For example, jargon aphasia is where a person produces incomprehensible utterances, such as [dɛlkwɑɪ] for 'lobster' and [ædɛpgʊd] for 'spade' (Rohrer et al. 2009). Zeigler and Maassen (2004: 427) observe that a striking feature of neologistic jargon is that it conforms to syllable structure conditions, suggesting that the problem is within the phonological module, or with the transmission of lexical information to the phonological module (Butterworth 1979). Similarly, some neurological conditions can affect lexical access and the phonological module. (e.g. schizophrenia can result in lack of realization of affix's phonological material – Cummings 2008: 365).

3.3 The PhM: Multiple modules and pooling

Generative theories allow for a single human to have more than one PhM. This situation is most obvious in the case of those who speak two different languages natively; they have distinct PhMs – one for each language (Genesee and Nicoladis 2007). For the same reason, people who speak distinct dialects of the same language natively arguably have more than one PhM.

An individual may also have several different speech styles – a particular way of talking depending on the social context (Labov 1972, Trudgill 1999). Some speech styles' phonological properties can differ in significant ways. For example, many Sāmoans have (at least) two speech styles: *tauta lelei* has the stops [p t k m n ŋ] while *tauta leaga* has [p k ʔ m ŋ] (Mosel and Hovdhaugen 1993: 20-24). In my own speech, I have styles that differ in /l/-vocalization, flapping environments, and declarative intonation tunes (Bye and de Lacy 2008§6). Distinct speech styles might be generated by different PhMs; alternatively, a single PhM could generate several speech styles if processes are indexed for speech style (see Coetzee and Pater 2011§4.5 for an overview of formal theories of speech style). The result is in effect the same as assuming multiple PhMs: each configuration of active processes (e.g. active rules/rankings) is a different state of the PhM, and the goal of phonological research is to determine the range of possible PhM states.

Taking languages, dialects, and speech styles into account, it is likely that there is no adult human who has only one PhM. So, it is not possible to assume that all the speech that comes from one human comes from the same PhM.

It is crucial to know whether a set of speech data was generated by a single PhM or is an amalgam of several PhMs' outputs. In the latter case, the resulting data might not reflect *any* possible state of the PhM.

For example, the data set [bə́ʊ] *bill*, [bə́.l-əŋ] *billing*, and [míː.tə] *metre* could be elicited from my speech. From these forms, one might conclude that I have a PhM that has (1) /l/-vocalization (/l/→[ʊ] outside of onsets), and (2) no flapping of /t/ (at least after long vowels). However, I have *no* PhM that has /l/-vocalization and lacks flapping. My informal speech style's PhM has /l/-vocalization and extensive flapping ([míːɾə], [hɔ́spəɾəʊ] *hospital*, cf. [hɔ́spətli] *hospitally* 'somewhat like a hospital'). My formal speech style's PhM has no /l/-vocalization and very restricted flapping (allowed only after short stressed vowels: [bʌ́ɾə] *butter* vs. [míːtə], [hɔ́spətɫ] *hospital*) (Bye and de Lacy 2008§6). The real danger here is that the PhM might not have *any* state that both has /l/-vocalization and lacks flapping. So, the use of data pooled from different PhMs has the potential to seriously derail the understanding of PhM capabilities.

So, it is an error to assume that all data from a single human was generated by the same PhM; such assumptions can lead to incorrect conclusions about possible PhM states. In practical terms, elicitation must be controlled for language, dialect, and speech style: techniques to keep the elicited data coming from an L1 module – i.e. track and prevent code-switching and style-shifting – are essential (e.g. Milroy 1987: 60ff, Milroy and Gordon 2003: ch.8; see Broselow, Chen and Huffman 1997: 52 for an example). EC do not report their elicitation techniques or whether they determined L1 status. In contrast, Echeverría (1964:20-21) states that the recording sessions were about 40 minutes long, and comment further:

(6) *Description of Scene-Setting in Echeverría (1964:20)*
"Estas reuniones se realizaron siempre en una atmosfera de confianza, y varias veces fueron precedidas de conversaciones informales relativas al tiempo, el trabajo particular del informante, etc. Pudímos notar que sentían orgullo por colaborar en este estudio de su lengua." (These meetings were always conducted in an atmosphere of trust, and usually were preceded by informal discussions about the weather, the participant's job, etc. We observed that they were keen to collaborate in this study of their language."

The steps taken were no doubt intended to reduce the subject's anxiety. However, it is not clear that they would have activated the subject's L1 Mapudungun vernacular module. The preliminary discussion was no doubt in Spanish, and using informal and friendly Spanish might not have prompted the subject to use vernacular *Mapudungun* – formal Mapudungun might still have been used. Labov (1984) and Wolfram (2011) provide overviews of techniques for controlling speech style and style shifting.

EC do not discuss whether code-switching between Spanish and Mapudungun occurred. The subjects were bilinguals and the elicitation was done in Spanish (Echeverría 1964:19,20), so code-switching and code-mixing could have been frequent, and may have influenced the prosody on Mapudungun words.[4] There is also no mention of how speech styles were controlled, or which speech style the data belonged to. For example, there is no information about the scene-setting that was given to the subjects.

EC do state that five subjects contributed to the project. Presumably their data was pooled (there is no indication of who contributed which word). So, the reported data almost certainly came from several different PhMs, all of which could have had different stress patterns. Quite apart from idiolectal variation (i.e. variation in PhMs across individuals which were mutually intelligible), the subjects could have been speaking different dialects. Given Mariá's rejection of Ailío's speech discussed in the previous section, it is possible that Ailío spoke a different dialect from the others.

Even if the speakers used only their Mapudungun PhMs and – conservatively – each had three speech styles, then there is the chance that EC's description is the pooled data from 15 PhMs. So any theory of EC's data would not be about a state of the PhM, but rather about pooled data from many PhMs, which may not bear any resemblance to any possible PhM state.

In practical terms, one might think that the preceding comments are overly critical. After all, EC's description is remarkably coherent in that it seems to be a single unified system. Such coherence might reasonably be seen as indicating that the subjects all had PhMs that had identical metrical systems.

The best way to test this assumption would be to compare the same set of words from all the speakers. However, EC gives no indication of who contributed which word (nor does Echeverría 1964). Going from EC's description it is even possible that the coherence is illusory: it could be the case that every speaker contributed different data, with no overlap. An example is found in Pankratz and Pike's (1967:fn.1) description of Ayutla Mixtec where the segmental phonology data and prosody data were elicited from different subjects. Echeverría (1964:20) provides further information, however: all speakers were asked to say two Swadesh word lists, were asked to say short sentences, and also told short stories. This elicitation procedure would

[4] For example, Gu et al. (2008) report that in Cantonese-English code-mixed speech, the prosody of the embedded language (English) was significantly influenced by the matrix language (Cantonese).

allow for direct comparison between speakers, so differences could be noted in pronunciation. However, there is little indication by EC that such a comparison was undertaken.

Echeverría (1964:47-8) also discusses another method employed: Mapudungun sentences were read to one subject (María Catrileo) with varying stress. She was asked to say whether the pronunciation was acceptable or not. If María Catrileo's responses were the sole basis for determining the stress system by EC as well as Echeverría (1964), then the results come closer to being useful for Generative analysis: the results were not pooled, and so likely to have come from a single L1 given María's native speaker status and frequent use of the language.

If the speakers had different PhMs, then some apparently 'free variation' in the data might be expected. In fact, EC report a great deal of allophonic variability. For example, "in final unstressed position, the nuclei /u/ and /o/ alternate" (EC 134). It is not clear whether this variation is seen within an individual's speech style, or whether /u/ is used in one speech style and /o/ in another, or one subject used /u/ and another /o/. Similarly, the word for 'red' is listed as both /kélï/ and /kelḯ/; there is no indication whether these words show free variation, speech style difference, or cross-subject variation. Similarly, a "non-phonemic glide" may optionally appear word-initially before a central vowel (EC 132), and much allophony involves free variation (EC 133). Again, there is no way of knowing from the description whether the variation is truly free – i.e. varies within a single PhM – or that different PhMs generate different allophones.

Due to data pooling, there is even no way to know how to interpret the variability in stress reported for CVCV words. It is possible that one speaker has consistently initial stress, another had consistently final stress, another had free variation, and another had lexically-determined stress. All possibilities fit the data provided.

To approach the issue from another angle, EC state that "2. Two-syllable words ending in a vowel may be stressed on either syllable … /pu rúka/ ~ /pú ruká/ *in the house*". A variety of PhM configurations could result in this data. Let us call a PhM which only produces /pu rúka/ 'P1', a PhM which only produces /pú ruká/ 'P2', and a PhM with free variation between /pu rúka/ and /pú ruká/ 'P3'. There are five configurations of PhMs which would fit with EC's statement: i.e. (1) P3 alone, (2) P1 and P2, (3) P1 and P3, (4) P2 and P3, and (5) P1 and P2 and P3. So, data pooling leads to significant uncertainty about which PhM states actually existed – the best conclusions to be drawn from the data are that (1) at least one of P1, P2, and P3 occurred, and (2) if P1 or P2 occurred then so did some other PhM state.

Finally, the relative coherence of the account may be illusory. The aim of the article seems to be to provide an account of sound patterns in the Mapudungun language. 'Language' is a socio-political concept, with no particular relevance to Generative phonology; for Generativist phonologists, the object of study is the PhM. The desire to describe a 'language' leads naturally to quashing variability: a language is a shared system of communication, so any language description is an account of normative behavior rather than of individual variation. In other words, the description might have aimed to identify general tendencies across all the speakers, and so omitted significant individual deviations.

One could argue that the validity of EC's description receives support from other descriptions that are very similar to EC's. Gordon (2002: 522) cites two other languages with left-to-right quantity-sensitive iambs, so one could assume that EC's description is *more likely* to be accurate than not. The problem with this justification is that it assumes that there is strong evidence for the other descriptions.

To summarize, pooling data generated by different PhMs into one dataset is potentially irrelevant and at worst misleading. If by chance the PhMs are identical in the relevant way, then the pooled dataset has no effect on the validity of the conclusions one can draw about the possible states of the human PhM. If, however, the PhMs are not identical, then the resulting dataset may not represent any possible state of the PhM, and theories of the PhM that rely on such evidence are not valid.

Data pooling in descriptions is ubiquitous. There are three reasons for pooling. One is that the aim of many descriptions is to describe a ‘language’: a set of generalizations about a group’s communicative behavior (see further discussion in section 4). For this goal, pooling is necessary: it is the only way to identify group norms. Another reason is that the goal of an article can be to describe normative behavior rather than states of a cognitive module. In contrast, Generative phonologists use data to identify states of the PhM, so pooling data from different PhMs can lead to the problems outlined above.

The third reason for pooling can be to minimize individual (random) variation. There are so many incidental influences on an individual’s speech that irrelevant ‘noise’ might swamp generalizations. By pooling several people’s data, it is hoped that the contribution of the Competence mechanisms will shine through. The problem with this approach is that it makes an a priori assumption that the speakers all have the same PhMs, or at least the same subparts of PhMs. Although the issues are complex, a more effective approach is to minimize individual variation by ensuring that data is elicited from the same PhM in an individual (i.e. controlling for code-switching and style-shifting), and that adequate amounts of data are elicited by using many different methods.

3.4 The PhM: Process interaction

Generative theories of phonology have stress-sensitive phonological processes.[5] Typical stress-sensitive processes include vowel reduction (e.g. Crosswhite 2004), fortition (e.g. Bye and de Lacy 2008), and allophony (e.g. Beckman 1998). So, such processes can be a source of evidence for stress position, a point made for EC’s description by Hayes (1995: 268). In fact, in the face of pervasive lexical stress, Altshuler (2009) crucially relied on phonological processes to determine the default pattern in Osage.

EC identify stress-sensitive allophonic processes:

(7) *Stress-sensitive allophony*

(a) */ï/*

“The phoneme /ï/ has a high back unrounded allophone [ï] in stressed position, e.g. /kelḯ/ [k̪̂elḯ] *red*, and a mid central unrounded allophone [ə] in unstressed position, e.g., /ṭómï/ [ṭʳɔ́mə] *cloud*.” (EC 132)

(b) /u/~/o/

“In final unstressed position, the nuclei /u/ and /o/ alternate: /fílu/ ~ /fílo/ *snake*.” (EC 134)

[5] Here, ‘stress’ in a phonological context refers to phonological elements that are realized as phonetic stress. Such elements include gridmarks (Prince 1983), and foot heads (e.g. Hayes 1981).

(c) Intonation
"Intonation is predictable from stress … and juncture." (EC 134)

Of course, it is necessary to put the claims in (7) through the same rigorous evaluation as the stress system. For present purposes let us assume that they are all valid and reflect active phonological processes. In practical terms, the allophony of /ï/ has the potential for independent evidence for the position of stress in 20 of EC's 72 forms marked with stress. For example, the word /mïlṭ ḯn/ *a kind of bread* would presumably be realized as [məlṱɯ́n], not *[mɯlṱɯ́n], *[mɯlṱə́n], or *[məlṱə́n]. EC provide 12 surface forms that show the [ə]~[ï] allophones. However, the description mostly consists of non-surface forms, so the allophones of /ï/ are not shown. Consequently, though the allophony is a potential source of evidence, the evidence itself – i.e. the allophonic transcriptions of forms – is not provided systematically in the description.

EC mention an apparently related process where "phonetically complex onsets … alternate with [CəC] sequences, and are analyzed phonemically as /CïC/, e.g. /kïθáw/ *work*: [kəθáu̯] ~ [kθáu̯]" (EC 134). As a diagnostic for stress position, this process does not add to the observation from /ï/ allophony that [ə] inhabits unstressed syllables. Smeets (1989: 63) considers this process epenthesis, and observes that insertion of [ə] is not sensitive to stress position – the inserted schwa can receive the stress.

The same point can be made for (5b): the /u/~/o/ alternation. I interpret the statement as meaning that /u/ and /o/ are in free variation in unstressed final position. Therefore, wherever final /u/ and /o/ do not freely vary, they must be stressed. This process would be useful in identifying stress position in six of the forms that EC cite (e.g. /elúmuyù/ *give us* should not vary with */elúmuyò/; /ṭipánto/ *year* should vary with /ṭipántu/). However, no comment is made on /o/~/u/ variants of each form. For example, *woman* is listed as /θómo/, and no /θómu/ variant is mentioned; no /o/-final variant of /sánču/ *pig* is listed.

EC also identify intonation as being sensitive to stress position. Interpreting their description in autosegmental terms, a phonological phrase has the tune LH*L where H* attaches to the leftmost (primary) stressed syllable and spreads to the rightmost primary stressed syllable in a Prosodic Word; moras may only bear one tone. For example, [{mànˈsánà}{kɔ̀ˈšílù}] *sweet apple*, [{wùˈlé}{màˈwḯˌnài}] *it will rain tomorrow*, [{tə̀ɸˈéítˢí ṭʳɔ́ˈmḯ}] *that cloud?* (my interpretation of data from EC 134-135). So, an initial L tone is diagnostic of whether the first syllable is stressed or not, potentially distinguishing the stress in words such as /píči/ *small* (i.e. as tonally [ˈσ́σ̀]) from /ŋïrḯ/ *wolf* (tonally [σ̀ˈσ́]). However, not enough intonation transcriptions are provided to demonstrate that the stress position on other data matches with the tones.

Overall, it is not the case that for every form cited there is corroborating evidence from /ï/ allophony, /u/~/o/ variation, or L tone placement to determine whether the transcription accurately reflects the location of phonological stress. So, while the processes offer potential diagnostics for stress position, EC does not contain systematic data that exhibits these processes, so there is little evidence to support the claims about stress position.

Much more can be said about the PhM's internal processes. For example, many Generative phonological theories allow stress to be sensitive to a variety of phonological environments and properties (e.g. tone, sonority, moraic content). So, adequate evidence will include forms that show the (in)sensitivity of stress to those contexts and properties.

Similarly, word stress and its realization interact with phrase-level prominence and intonation. Gordon (this volume) makes a particularly pertinent observation: in stress descriptions, the reported patterns often appear to refer to words spoken in isolation. In such cases, words are coextensive with utterances so word stress would interact with phrase-level intonation and prominence. For some descriptions, it might be impossible to tease word-level stress and phrase-level properties apart, leading to an inaccurate understanding of prosodic structure at the word-level. EC do not state whether their stress description is based on words spoken in isolation or in phrases. So, EC's assertion that stress falls on the second syllable in /putún/ *to drink* could be an observation about word-level stress, phrase-level prominence, intonation, or the interaction between these factors. Even if the words were extracted from a larger context, a variety of challenges arise: the realization of word stress can differ based on phrasal position, phrase-level prominence can be sensitive to factors that are different to word-level stress, and intonation might not match well with word-level stress. Lack of space prevents further discussion of this point here; the reader is referred to Gordon (this volume) for extensive discussion.

The broader point is that the phonetic manifestation of foot heads is not the only evidence for phonological stress. If word-heads, foot-heads, and foot boundaries are present in the phonological representation, phonological processes can refer to them. So, such processes provide extremely important support for the location of phonological stress. An important caveat is that stress-referring phenomena may not always occur in the presence or absence of stress. For example, Hyde (this volume) observes that descriptions of Tauya have claimed that the initial syllable is stressed because it resists vowel reduction. However, Hyde suggests that the lack of reduction could be due to special restrictions on word-initial syllables.

3.5 The Phonetic module

In the production process, the PhM's output goes to the Phonetic Module, and is converted there into a gradient representation that is ultimately translated into articulator movement (Keating 1990, Kingston 2007§17.4.3). There are many different conceptions of how the PhM and Phonetic Module interact (Löfquist 1997, Kirchner 1998). For present purposes, what matters is that the translation of phonological forms into instructions for articulatory movements can introduce distortion.

The Phonetic Module is not straightforward: it is not the case that every phonological element has a realization that is invariant across speakers, or even for one speaker in different contexts (Kingston and Diehl 1994). In fact, a variety of work has shown that phonetic processes can obscure phonological information. For example, anticipatory coarticulation can make a phonological [np] be realized as (something very close to) a phonetic [mp] (e.g. Browman and Goldstein 1995). The Phonetic Module also adds structure, as in tonal interpolation (Gussenhoven 2007§11.3) or featural interpolation (e.g. Choi 1995). It can also displace phonological symbols by realizing them out of temporal sequence (e.g. glottal stops realized as creaky voice in Pendau – Quick 2004; displacement of pitch accents – Gussenhoven 2007§11.4.2). Finally, some phonological symbols have no direct phonetic realization (e.g. Cairene Arabic secondary stress – Allen 1975).

Phonological stress is particularly interesting because it has a variety of realizations. For example, stressed vowels can be realized with an excursion in fundamental frequency (F_0), increased intensity, lack of spectral tilt, increased duration, vocalic peripheralization (or centralization in unstressed syllables), a combination of the preceding, or no realization at all (e.g. Lea 1973: 27ff, de Jong 1995, Gordon 2011 and references cited therein).

EC do not describe the phonetic realization of stress.

A problem with not knowing the realization of stress is that the perceived stress could be an artefact of the analysts' perceptual systems. For example, if an analyst has a perceptual system in which the realization of stress involves raised pitch, then the analyst may perceive raised pitch in the target language as indicating the position of stress. However, the speaker's stressed syllables may not be marked significantly by F_0; instead, F_0 may be controlled by incidental phonological factors (e.g. intonation) or phonetic ones (e.g. a rise in F_0 word-initially).

Without a full analysis of how the subjects' Phonetic Modules realize phonological stress, it is not possible to be confident that EC's description of stress position is accurate. Nowadays, techniques to help avoid such difficulties are in common use (e.g. machine recording and sound analysis – Ladefoged 2003). Of course, such techniques were not widely available in 1965. In fact, Echeverría (1964:21) noted that spectrograms would have been very useful for transcribing, and that some aspects of Mapudungun speech were difficult to perceive (particularly dentals and retroflexes). However, the point remains that without a clear characterization of the phonetic realization of stress, and clear evidence that this realization does in fact occur where EC transcribed stress symbols, it is possible that EC misperceived stress.

Another difficulty with stress is that other phonetic processes can obscure it. For example, F_0 is often higher phrase-initially than phrase-finally due to a possibly universal tendency for F_0 to decline over a phrase (Ohala 1978: 31-32); this natural declination could be misinterpreted as initial stress. Similarly, a variety of phonetic processes cause changes in vowel length, such as pre-voiced consonant lengthening (Chen 1970), domain-initial strengthening (Cho and Keating 2009), and domain-final lengthening (Beckman and Edwards 1990). Where vowel length is an important cue to stress position for the analyst but not in the target language, such incidental duration changes could be misinterpreted as stress.

O'Keefe (2007) reports a pilot study of Mapudungun CVCV words to determine whether there were clear acoustic correlates of stress. Using limited data from audio recordings on the compact disc accompanying Zúñiga (2006), O'Keefe did not find significant differences between syllables in CVCV words: intensity was not statistically significant (n=13, two-tail t-test: p=0.198), nor was F_0 (n=13, two-tail t-test: p=0.064). There was a significant difference in duration: final syllables tended to be longer than initial syllables (V1 mean=76ms, V2 mean=97ms; $n=13$, one-tail t-test: p<0.01). However, the average difference of 21ms is small and its perceptual significance is questionable; it is possibly due to domain-final lengthening. Even though O'Keefe (2007) did not work with EC's subjects, the uncertain results of his work suggest that identifying the correlates of stress among Mapudungun speakers is not straightforward.

In summary, without a clear understanding of the phonetic realization of stress, and of other interacting phonetic processes, there is an added element of uncertainty in EC's description. It is possible that what EC describe is an artefact of their perceptual systems, or a misinterpretation of incidental effects on F_0 as indicating stress position.

The preceding discussion might be read to imply that the only descriptions of stress acceptable for Generative analysis are those that are accompanied by detailed analysis of phonetic implementation (e.g. Gordon 2004, Altshuler 2009). There has certainly been a significant rise in such work over the past several years (Cohn 2007). The issue of how the observations made here relates to the notion of 'acceptable evidence' is discussed in section 4.

3.6 The neuro-motor interface and articulators

The neuro-motor interface implements phonetic representations as articulator movement, and can be affected by a variety of disorders (Weismer 1997). Similarly, inherent properties of articulators can distort the PhM's output, and so can articulatory disorders (e.g. Bernthal 2009).

Some articulatory-sourced distortions are found in all humans. Articulators cannot move as quickly as they are commanded in all situations. So, the shorter duration of unstressed syllables can mean that articulatory targets cannot be met in time, leading to vowel centralization (Lindblom 1963). Such centralization can create an ambiguity for the hearer/analyst: vowel centralization might be evidence for either phonetic centralization or phonological neutralization (e.g. Barnes 2002).

Certain articulations have inadvertent side-effects. For example, jaw and tongue height can inadvertently influence pitch, so that high vowels have higher F_0 than low vowels (Whalen and Levitt 1995); there are many other similar influences on F_0 (Lea 1973). So, even though the PhM can also influence F_0 tonal specifications and stress, an F_0 excursion is not unambiguous evidence for a PhM specification.

It is a challenge for the analyst to tease apart the effects of articulatory implementation from the phonetic representation (and then tease apart the phonetic and phonological representations). For example, jaw and tongue height affect duration, with lower vowels having longer inherent duration than higher vowels (Lehiste 1970). If duration is perceived as the cue to stress position, low vowels might be misinterpreted as having phonological stress. There is the chance that this particular articulatory effect might have led to misanalysis of inherent vowel duration as phonological stress – a possibility not considered in the most extensive surveys of sonority-driven stress (e.g. de Lacy 2004).

Articulators can vary a great deal between humans. For example, laryngeal size differs significantly based on age and sex (Beck 2007). An effect is that average F_0 for a 10-year-old male or female is around 260Hz, for a 20-year-old female it is around 200Hz, and a 20-year-old male has around 130Hz. Apart from average F_0, speakers apparently often differ in pitch range. F_0 is often significant for the detection of stress. If stress is realized with an F_0 excursion, the nature of that excursion may be significantly different for a speaker with a large pitch range compared with a speaker with a compressed pitch range. Similarly, pitch rises might be difficult to detect in speakers who have a higher average F_0 as their pitch may reach its peak without much deviation from the average.

Trauma, medication, and disease can also cause articulator and neuro-motor variation (e.g. Weismer 1997). The sources of variation may be due to age degeneration or to a specific traumatic/disease event, and may be permanent or transitory. For example, a common age-related degeneration is loss of all natural teeth (edentulism). Edentulism can affect the production of dental sounds (e.g. [v], [θ]). In the United States, its prevalence is 2.3% for 18 to 44 year olds (all races/ethnicities and incomes), and 25.3% for 65 year olds and older (age-adjusted).[6] The highest rate is among 85 year olds and older who are also poor and Hispanic (male and female data pooled) at 66.1%.

[6] US Centers for Disease Control and Prevention (CDC): Health Data Interactive: Edentulism (loss of all natural teeth, ages 18+: US, 1998-2009 (Source: NHIS)

Alcohol is an example of a transitory influence: Chin and Pisoni (1997) provide an extensive discussion of research into the influence of alcohol on speech production and perception. They note that for one study, the "most consistent acoustic change was a significant increase in sentence durations, that is, a slowing of speaking rate under alcohol. In addition, control of abrupt closure and opening of the vocal tract were also affected by alcohol" (p.184). They also observed increase pitch variability and impairments of fine articulations involving articulator timing.

With alcohol, the Phonological and Phonetic modules are (presumably) intact, as are the articulators. The impairment seems to center on the neuro-motor interface, resulting in mistiming of articulatory commands. Chin and Pisoni (1997: 185) identify problems with the production of voiced stops, affricates, and stop clusters, all of which require fine control of articulator timing.

Alcohol therefore poses potential problems in the analysis of stress. If pitch alignment and duration provide cues to stress position and alcohol intoxication interferes with the accurate timing of articulatory movement, the result could be that an F_0 excursion occurs in a place unrelated to where the PhM specifies it should be.

There are many ways in which the neuro-motor and articulatory mechanism can distort the intended realization of the PhM's output. For present purposes, the point is that impairments of the neuro-motor interface can make recovery of the PhM's output difficult, and even obscure it entirely (Weismer 2007; e.g. dysarthria – Bernthal 2009: 172ff; soft palate implants – Akpinar et al. 2010).

The practical implication is that certain physical properties of the data sources (i.e. the subjects) must be reported because they can distort cues to stress, leading to misanalysis. Sex and age are commonly reported; these properties at least give an indirect sense of potential characteristics of the speaker's articulators (Beck 2007). It is also apparently necessary to examine individual variation and impairment in at least the articulatory apparatus. If the elicitation procedures involve visual and/or auditory stimuli, then visual and/or hearing impairments may also need to be reported, depending on what is being tested.

EC do not report any of the characteristics of their subjects. They refer to them as "five informants", with no indication of sex, age, or impairments. They do not give the subjects' names so it is impossible to infer sex. Consequently, there is an additional uncertainty with the data EC provide: it is not possible to determine how the articulatory variation among EC's subjects influenced EC's perception of the location of stress. In contrast, Echeverría (1964:19-20) provides some relevant information: the subjects were three males and two females, with ages 18, 32, 35, 40, and 50. No information about the subjects' health is provided.

It is reasonable to be concerned about articulatory variation and impairment in the Mapudungun context. In 2008, the Mapuche were reported to "suffer from poor housing, malnutrition, illiteracy, alcoholism, tuberculosis, and a high rate of infant mortality." (UNPO 2008). It is possible that the same situation existed in 1965, or at least that there were precursors to the problem.

Rates of alcohol abuse and use in many populations are significant (e.g. US: Nace 2005; Australia: AIHW 2007). As alcohol's effects on speech and perception are well established, use of valid tools for evaluating subjects for alcoholism and alcohol intake may well be essential (e.g. Selzer 1971, Babor et al. 2001).

---

http://205.207.175.93/HDI/TableViewer/tableView.aspx?ReportId=110 [retrieved Feb 6, 2012]. The percentages are from table values as specified in the text.

Alcohol was chosen arbitrarily here as merely one source of variation in an individual's module. The mere fact that an individual does suffer from alcoholism or acute intoxication of course does not automatically disqualify them as sources of phonological evidence. However, it introduces effects on the perceived speech sounds that must be factored out when determining the PhM's contribution. There are many kinds of individual variation involving articulators. Even in relatively wealthy, non-marginalized populations, there are often high rates of impairments that can potentially affect speech production. For example, oral pain caused by tooth decay could influence speech production. In the USA 23.1% of all people between the ages of 20 and 44 had untreated tooth decay (CDC 2008); in socio-economically marginalized groups the rate was much higher (e.g. 'Poor': 39.8%; 'Poor Non-Hispanic Black': 48.7%).

In fact, it is unclear how many humans *lack* impairment (whether acute or chronic); perhaps they are a small minority of the population. The implication is that if a phonological description does not report impairments of the speaker (or lack of them) there is necessarily uncertainty about whether the analyst's interpretation attended to properties of the speech signal that were controlled by PhM specifications, rather than being caused by neuro-motor or articulatory properties.

3.7 The perceptual process

The preceding sections have focused mainly on the speaker. However, most grammars and articles report what analysts perceive the speaker to have produced, so the hearer is an important source of distortion.

The speaker's articulators cause variation in air pressure and flow. However, the subject's articulators are not the only potential sound source. Other speech and machinery can create noise that can distort perception of the speech signal. To minimize such noise pollution, laboratory experiments can use sound-proof booths or enclosures, and fieldwork can be located in quiet rooms. EC do not report where they elicited data, so it is unknown whether background noise interfered with accurate transcription of some data. Similarly, there is no information about EC's perceptual systems, assuming that their transcription is entirely impressionistic.

Just as with production, each module and mechanism in the perception process provides an opportunity to distort the speech signal. Damage to the ears and hearing mechanism can mean that certain phonetic cues cannot be observed, or that other cues are introduced (and damage is more likely with older people: e.g. Pinchora-Fuller and Souza 2003). For example, partial hearing damage can prevent discrimination among fricatives (Zeng and Turner 1990) and many other contrasts (e.g. Boothroyd 1984).

Perhaps the most important issue is that learning an L1 involves a person's L1 perceptual mechanisms losing the ability to discriminate between certain sounds that do not have a contrastive function (Werker and Tees 1984; Best 1999). Consequently, it may not be possible for a person to reliably discriminate sounds that they do not need to distinguish in their own speech. For example, Peperkamp & Dupoux (2002) report cases where people have 'stress-deafness' – an inability to perceive stress distinctions due to the low functional load stress carries in their L1.

As mentioned in section 3.5, stress can manifest itself in many ways. So, the cues an analyst has learned to attend to may not be present or relevant in the subject's speech. Without knowing details about EC's perceptual systems, it is not possible to know which phonetic cues they were attending to, and which they were ignoring.

In fact, Hayes (1995) uses the uncertainty surrounding EC's perceptual systems to cast doubt on certain aspects of their description: "I assume that the prominence heard on final CVC syllables by Echeverria and Contreras is a perceptual effect" (Hayes 1995: 266). (In contrast, Kager 1993: 409 assumes that EC's description of final CVC stress accurately reflects the speaker's intentions.) In other words, EC's perceptual system might have attended to phonetic cues that were not diagnostic of the subject's placement of stress. Hayes (1995)'s theory does not allow for CVC syllables to be heavy in this particular case, so the uncertainty about EC's perception raises the possibility of an analysis that differs from their description. Of course, one could observe that most aspects of EC's description could be called into question in the same way: perhaps *all* secondary stress reported was misperceived.

A number of problems with impressionistic transcriptions have become apparent (e.g. in the discovery of incomplete neutralization – Kawahara 2011§2.2.2 and references cited therein). There are a variety of techniques for avoiding them, involving machine recording and software. However, recording devices and analytical software pose issues analogous to impressionistic transcription. Like ears, microphones are imperfect, and may distort the incoming signal due to limitations relating to directionality, impedance, frequency range, and frequency response. For example, microphones with high impedance will lose high frequencies when used with long cables. All microphones exaggerate some frequencies and attenuate others (frequency response); microphones also differ in which frequencies they can pick up.[7] Like the neuro-auditory interface, digitation mechanisms (e.g. computer sound cards) may distort the signal due to limitations on sampling and frequency range and response. Like the human perceptual mechanism, software generates data structures that can distort the signal, such as pitch trackers inadvertently halving or doubling F_0 (Gussenhoven 2004:5-7). Finally, there is still a human element: the analyst must decide what to attend to and what to ignore in the spectrogram, pitch track, or other measure. Echeverría (1964:20-21) reports that the elicitation sessions were recorded and the recordings were used in transcription. However, no information about the recording apparatus is provided.

What makes using machines and software potentially better than humans is that it is possible to know exactly how machines distort the incoming signal, while the human perceptual system is still mysterious. However, just like people, machines have a great deal of individual variation. So, without careful specification of which machines, software packages, and analytical methods were used, along with careful attention to their limitations, machine-derived data may be of as little value as impressionistic transcriptions that lack information about the analyst's perceptual system.

Molineaux (2012)'s work on the perception of Araucanian/Mapudungun stress is revealing here. Molineaux played two-syllable Mapudungun words to several *non*-native speakers and required them to identify primary stress. For CVCV words, a little over 30% reported final stress, while around 50% reported initial stress (the remainder did not respond). Such disagreement among different non-native listeners might indicate that the perceptual cues for stress in Mapudungun are quite different, or not as robust, as those in the non-native speakers' languages.

Even so, impressionistic reports and intuitions cannot be dismissed lightly. Hualde and Nadeu (this volume) argue that "data from intuition are worthy of attention to the extent that they are consistent among speakers of the language." Even if intuitions are a delusion, in some cases they are a shared and consistent delusion, and the source of the delusion might provide insight

---

[7] http://www.mediacollege.com/audio/microphones/how-microphones-work.html provides an accessible tutorial.

into phonological knowledge. The problem for Mapudungun is that native speakers seem to not have consistent intuitions about stress: Molineaux (2012) reports that “native speakers were uncertain about stress in disyllables”.

Hualde and Nadeu also suggest that intuitions about contrastive (lexical) stress might be more reliable than non-contrastive stress. They observe that there is universal agreement about the placement of primary stress in Spanish but disagreement about secondary stress; primary stress is contrastive, secondary stress is not. However, even if there is universal agreement about some speech property, it is always a significant leap to claim that the property directly reflects a particular aspect of the phonological representation or derivation. I suspect that intuitions are complicated – they can be influenced by the perceptual, lexical, and orthographic systems, and so are unlikely to always directly reflect phonological knowledge.

3.8 Expression of the data

The final step for the analyst is expressing the perceived speech sounds (PSS) in written form. Obviously, the transformation from PSS to written word is hugely complex. Here, the focus is on the challenges of adapting a description expressed in one theory for use in another theory.

EC express their description in a taxonomic phonemic framework. They do not propose rules in the formalism proposed by Chomsky and Halle (1968), so there is no evidence that they had Generative theories in mind.

Translating from PSS to a phonemic representation is a complex task (Harris 1960). Principles of phonemic analysis are applied to determine contrastive sounds, allophonic relationships are identified and the more basic allophone is selected as the symbolic representation. So, /kúra/ *stone* is not a Generative underlying (or intermediate) form: it is a description of contrastive relationships between speech sounds, with stress also marked. Most of EC’s transcriptions are not the equivalent of Generative PhM output forms – they do not show allophones. However, EC do provide some allophonic forms, and even phonemic-allophonic pairs: e.g. /ṭómi/ and [ṭʳɔ́mə] *cloud*; /ïnáy/ and [ɣənáḭ] *he scratches himself* (EC 132).

The challenge for anyone who wishes to use EC’s description is to convert it into a form relevant to their theory. In doing so, it is necessary to carefully examine EC’s theoretical decisions to see if any information crucial for other theories was removed or altered.

For example, in classical single-level Optimality Theory (OT) candidates are PhM output forms (Prince and Smolensky 2004), and so are roughly equivalent to a phonemic analysis’s (broad) allophonic transcription. So, EC’s “/tómï/” the OT winning candidate would not be [t̪ómɯ], but rather the surface/allophonic form [t̪s̪ɔ́mə].

The pitfalls in using a phonemic description for output-oriented Generative theories are illustrated in EC’s description of five-syllable words. EC cite /elúaènew/ *he will give me*, and /elúafimi/ *you will give him* (presumably /elúafìmi/, with secondary stress omitted for expository purposes). However, these forms are phonemic representations. EC’s *allophonic* transcription of *eluafimi* is the *four*-syllable “[… elu̯ áfimì…]” (IPA [ɛlwáfimì]). If glide formation is a general process, then the allophonic form of /elúaènew/ would also have four syllables: [ɛlwáɛnɛ̀w]. So, there is a difference between EC’s description and its expression in OT: EC’s five-syllable phonemic forms must be translated into their four-syllable allophonic counterparts to be used in output-oriented theories. In fact, from the point of view of output-oriented theories

EC cite *no* five-syllable forms. (The only other five-syllable form /ḽaŋɨ́mïwulày/ *he did not kill himself* has a suffix that interferes with stress placement.)

A more profound challenge of interpretation involves quantity-sensitivity. EC's General Rule appears to describe a quantity-insensitive iambic system: alternating stress without regard to syllable shape. However, for EC stress is only quantity-insensitive *early* in the derivation. Later in the derivation certain rules have applied that mean that stress apparently becomes quantity-*sensitive*. For example, EC report the allophonic form "[wəṭʳéimi káḭ] *are you cold?*" (EC 135). The first word must be syllabically parsed as [wə.ṭʳéi.mi] otherwise stress would be distributed as *[wə.ṭʳé.i.mì]. The same point applies to [mawɨ́nài]: it must be syllabified as [ma.wɨ́.nài], not *[ma.wɨ́.na.ì] (EC 135). With an iambic footing analysis, quantity-*sensitive* iambs are necessary to get the right stress: [(ma.wɨ́)(nài)], not quantity-*insensitive* *[(ma.wɨ́)nai]. However, in EC's phonemic analysis there are no CVV syllables when stress is applied, and so no quantity-sensitivity. Of course, a surface-oriented OT analysis requires evaluation of *output* forms. So, to have [(ma.wɯ́)(nài)] beat *[(ma.wɯ́)(na.ì)], an analysis that treats CVV syllables as heavy is required. In short, while EC's phonemic theory allows them to treat stress as quantity-insensitive (at least in the early derivation), OT requires the same system to be analyzed as quantity-sensitive.

Of course, the preceding discussion assumes that EC's allophonic transcriptions accurately represent the PhM output form after all phonological processes have applied; they might in fact show derivationally intermediate forms, or describe Phonetic Module representations, or show the application of optional processes.

A thorough evaluation of a source must address the issue of what the source did not mention and why. A description's silence about a phonological property does not mean that the subjects failed to exhibit the property. A common example of inadvertent omission is secondary stress. Many descriptions do not mention secondary stress at all, even to deny that it is present: even casually browsing through the UD Phonology Lab Stress Pattern Database emphasizes this point – many entries note that the lack of secondary stress is "deliberately ambiguous between "none reported" in a source and "verifiably none"" (Heinz 2012). Gordon (2000) documents several cases of default-to-opposite stress systems where a source's silence about secondary stress was incorrectly interpreted to mean that it did not exist.

Secondary stress presents a particularly difficult problem because it seems to be far more difficult to perceive than primary stress. If so, primary stress descriptions might be more reliable than secondary stress descriptions. (It is also possible that rhythmic secondary stress might be easier to perceive than non-rhythmic secondary stress, such as secondary stress on initial or final syllables). EC did report secondary stress; however, without knowing how secondary stress is phonetically realized or how other phonological processes are affected by it, we cannot be sure that EC correctly reported all secondary stresses, or were even correct in identifying the secondary stress they did. Notably Echeverría (1964) did not distinguish primary from secondary stress in his transcriptions.

In some cases, omissions are deliberate and are driven by theoretical concerns. For example, a stress theory that does not allow for more than two word-stress distinctions (primary vs. secondary) could lead to describers not reporting tertiary stress. EC's description is an interesting case because the theoretical goal for stress seems to have been to provide a rule for it rather than account for the distribution of stress in surface forms. Since the rule appears derivationally early in their analysis, later rules that significantly obscured the earlier assignment of stress (e.g. glide formation) were omitted from the description, along with the actually attested surface form of words.

The general point is that converting a description into a form that is useable for a Generative analysis is not a trivial task. If a source is expressed in a different framework (e.g. taxonomic phonemics), then repurposing it for a Generative analysis can be fraught with danger. In fact, if a source does not supply output forms, then it has limited worth for output-oriented theories: even when there is explicit description of phonological processes, lack of output representations introduces uncertainty as to whether all relevant phonological processes have been described. For example, EC does not mention glide formation or the processes that lead to vocalic diphthongs; their existence was deduced from their allophonic transcriptions. So, from the point of view of an OT analysis, EC does not provide the right kind of data: their theoretical view transmutes the data in a way that it is difficult to derive PhM outputs – i.e. the kind of data that output-oriented Generative theories require.

Hyman (this volume) notes that descriptions and analyses are theory-dependent; there is no way to present a theory-neutral description of stress. So, EC did not make a mistake by expressing their description in phonemic theory; it was inevitable that it must be expressed in *some* theory. In fact, recognizing the role of theories in description is valuable: it allows us to squarely face the issue of translating the description's theory into a Generative one. Of course, some (perhaps all) theories end up expressing the data in such a way that properties that are crucial for other theories might be obscured. For example, if one wishes to express a stress description in a theory that has culminativity, then necessarily one of the stresses in a domain must be described as the 'primary' one, even if there is no evidence for that claim. The description then becomes misleading for use in a theory without culminativity. In the most extreme cases, it might not be possible to adapt a description expressed in one theory for use in another theory.

3.9 Replication

Replication of stress descriptions is not required by any aspect of the theory. However, given the opportunity for errors at so many stages in the evidence gathering and expression process, replication can provide a way to increase confidence in the results. In the field of psychology, the issue of replication and the consequences of lack of replication have received a great deal of attention recently: Hartshorne and Schachner (2012) advocate tracking experiment replication; Sutton (2012) is a journal issue devoted to replication; and there is a list of 20 psychological experiments that people would most like to see replicated at http://psychfiledrawer.org/top-20/.

Data replication for the Generative phonologist would involve eliciting the same forms from the same PhM. So, for every elicitation/experiment involving a person's PhM, the same elicitation/experiment would have to be performed on that person's PhM. Less ideally, the same elicitation/experiment could be performed on a different person, as long as a strong case is made that the individual had a PhM that was identical to the original subject's in relevant respects. The replications should also not be greatly separated in time otherwise the subject's PhM might change, or other production mechanisms might degrade significantly.

A far less ideal data replication would be for the same elicitation/experiment to be performed on a subject who speaks the same "language" as the original subject. It is a less ideal situation because "language" is an irrelevant concept for Generative theories: it is a sociopolitical idea with a tenuous relationship to the object of study: i.e. the linguistic cognitive modules. The group of people who say that they speak the same language can have vastly different PhMs, even to the point of mutual unintelligibility. So, words elicited from a Mapudungun speaker in 1965 cannot be assumed to have any cognitive relevance for the same words elicited from a Mapudungun speaker now. In this sense, then, EC's description cannot be replicated – if they used the same subjects as Echeverría (1964), it is likely that some of them will now be deceased. For the others, they will be significantly older, and their Mapudungun PhMs may have changed or not been used over the past several decades. In fact, it is possible that too much time has elapsed to be confident that there is any person with a PhM that is identical to the original speakers.

However, as potentially problematic as replications involving different subjects can be, for stress descriptions they are the only type of replications available, so let us consider them valid replications for the sake of this discussion.

Stress descriptions used in metrical work are typically unreplicated. Of Gordon (1999)'s list of 392 metrical systems (i.e. 'languages'), 272 (69%) were listed with only one source and therefore unreplicated (assuming that Gordon 1999's bibliography is comprehensive). Of the remaining 120 languages, it was unclear how many of the later descriptions involved re-elicitation of the same data as the original description – the number of replicated descriptions may therefore be much lower than 31%.

Mapudungun is in the minority: there have been several recent descriptions of its stress patterns, including Echeverría (1964), and the grammars of Smeets (1989), Salas (1992), and Zúñiga (2000, 2006).[8] EC was published in 1965 whereas the other grammars came decades later, so it is highly unlikely they had the same subjects as EC. It is therefore quite possible that all the grammars describe distinct PhMs, in which case considering them to be replications of EC's description is incorrect, and comparing them would reveal nothing of value. With this caveat in mind, let us assume that the grammars' subjects did have identical PhMs.

There is a great deal of disagreement among the four descriptions. For example, for CVCV words, EC describe lexical stress (or free variation), Smeets (1989: 60) a tendency for initial stress, Salas (1992: 83) free variation, and Zúñiga (2006: 64) a tendency for final stress. For main stress, EC describes it as falling on the second (peninitial) syllable, while Salas and Zúñiga have it fall on the final syllable if it is closed, otherwise the penult; Smeets locates it on the second syllable, except in five-syllable words where it falls on the penult, and describes long words as having two primary stresses.

For secondary stress in polysyllables, EC have secondary stress on every even-numbered syllable after the main stress (e.g. /kimúfalùwulày/ *he pretended not to know*). Smeets says that every second and every final syllable has secondary stress (e.g. [allkɨ́penuèl] 'unheard of' – Smeets 1989: 60). Zúñiga (2006: 64) describes secondary stress as falling on the first or second syllable, depending on which one is closed (e.g. *kàmapuléy* 'it is far', *weyùlküléy* 'it is swimming').

Trisyllables of the form CV.CV.CVC illustrate the extent of disagreement well: EC CV.CV́.CV̀C (e.g. /θuŋúlàn/ *I do not speak*), Smeets CV.CV́.CVC (e.g. [kuñífal] 'orphan' – p.60),

[8] Augusta (1903:2-4) is so far removed in time from the recent grammars it will not be discussed here (see Echeverría 1964:46ff for comments).

Zúñiga CV.CV.CV́C (e.g. *machitún* 'healing ceremony' – p.64), and Salas CV̀.CV.CV́C (e.g. *à-cha-wáll* 'chicken' – p.84).

There are some similarities. For example, all sources agree that CVC syllables attract stress in some situations: e.g. all agree that stress is final in CV.CVC words.

It is quite possible (even likely) that all the sources were describing different PhMs, so the fact that they disagree is not proof that EC's description is inaccurate. However, at the very least the differences between the sources mean that EC's results have not been replicated.

An interesting comparison is Echeverría (1964) and EC since they apparently used some of the same recordings, and were therefore describing the same PhMs. However, there are analytical differences between the two descriptions. Echeverría (1964:48) states "El acento no es predecible fonemicamente, pero sí podría serlo en un plano morfémico, es decir, atendiendo a cierto tipo de morfemas." (Stress is not phonemically predictable, but could be on the morphemic plane – i.e. influenced by certain types of morphemes), and adds that certain morphemes probably have fixed accent (e.g. /rumé/ 'quite', /wulé/ 'tomorrow'). Echeverría (1964:47-48) also reports the result of a perceptual experiment where the speaker judged both [umau̯tulén] and [umáu̯tulén] to be acceptable; the speaker rejected forms with initial and third syllable stress. It is surprising that [umau̯tulén] was accepted given that EC describe primary stress as falling on the second syllable (p. 134.c2) (cf. /elúmuyù/ *give us* – 134.c2). In fact, Echeverría (1964) makes no distinction in transcriptions between primary and secondary stress: e.g. /féytɯfá/ 'este'/this (p.54), /l̪aŋɯ́mɯ́n̪/ 'matar'/kill (p.55).

Echeverría (1964:54-55) also provides transcriptions of 107 words and 43 phrases, with stress marked. When comparing the data with EC's, the same general results emerge for disyllables: CVCV words have mostly initial stress, CVCVC words have final stress, CVCCVC and words tend towards final stress.

(8) *Disyllable stress in Echeverría (1964:51-54)*

Shape	Initial stress	Final stress	Both	Total
CV.CV	27 (96%)	1 (4%)	0	28
CV.CVC	0	32 (97%)	1 (3%)	33
CVC.CV	4 (57%)	2 (29%)	1 (14%)	7
CVC.CVC	0	6 (100%)	0	6
ALL	31 (46%)	41 (55%)	2	74

However, in trisyllabic words Echeverría (1964) has several disagreements with EC's description. EC state that CV.CV.CVC words have penultimate primary and final secondary stress (e.g. /θuŋúlàn/ 'I do not speak'). However, Echeverría (1964) marks stress on final syllables: /aliwén/ 'tree', /ɯnatún/ 'bite', /weyelɯ́n/ 'swim', /nɯṭamkán/ 'say', except for two words /l̪aŋɯ́mɯ́n̪/ 'kill', /umáwtún/ 'sleep' which have equal stress on both penult and final. In agreement with EC, the second syllable is stressed in /fɯtáŋma/ 'large' and /mawíθa/ 'mountain', but other stress patterns given are /fɯtáŋmá/ 'long' and /féytɯfá/ 'this'. Only one word longer than three syllables is given: /ṭɯlkéaliwén/ 'peel,crust', and this stress pattern does not fit with EC's description of alternating stress (i.e. it should be */ṭɯlkéalíwen/).

4. Levels of acceptability

The aim of the preceding section was to use the theories of Generative phonology and other modules and mechanisms to identify possible sources of distortion of the PhM output on its path from the speaker to the analyst's description. There are many more conditions and other modules and mechanisms that should be examined (e.g. limitations on memory and attention). However, this section discusses the more practical issues of levels of acceptability and responsibility for evidence evaluation.

Until phonological outputs can be detected directly, all evidence will have some potential for non-PhM distortion, and so carry some degree of uncertainty. So, the challenge is not to achieve perfect evidence, but to decide how much uncertainty can be tolerated.

There is an objective and a subjective aspect to evaluating evidence. The objective side is that theories of speech production and perception identify places where distortion of the speaker's PhM output occurs. So, no list of evidentiary requirements can be merely prescriptive; each requirement must be shown to derive from a principle of the theory. For example, requiring that the eye color of the consultant be reported cannot be justified by any relevant theoretical principle (I know of no research that links eye color to production or perception). In contrast, controlling for speech style shifts is crucial as the theory has different PhMs (or PhM states) for each different speech style.

The subjective side is determining an acceptable level of uncertainty for any data. At what point does uncertainty mean that a description cannot be used as evidence? Are there certain points that are non-negotiable? For example, does data pooling mean that the description cannot be used?

I am not aware of any current consensus about minimal requirements for evidence in the field of phonology. Perhaps lack of explicit discussion and publication of such standards is hampering – or even thwarting – progress in the field. After all, if this chapter's exploration of EC's description occasions any pause among metrical theorists, then similar issues may well arise for many other commonly cited sources. It may well be worthwhile to require publications to explicitly evaluate any source used as evidence, or start to accept articles for publication whose sole aim is to evaluate the degree of certainty for a source.

Certainly, whatever the standards are that currently exist or develop, they must be reviewed frequently and allowed to change. Standards of acceptability change over time as theories, methods, and technology develops. The dominant paradigm for description in 1965 involved phonemic analysis, with the goal of determining phonemic forms. Stress (and other phenomena) could be described as applying to phonemic forms, so there was often no theoretical motivation for providing extensive allophonic forms (just as in EC's description). In contrast, the focus of Generative theories involves identifying transformations from underlying to surface forms, so descriptions must provide underlying forms, transformations, and surface forms (e.g. compare EC with Parker 1999). Particularly for descriptions expressed in output-oriented theories, the phonological output must be provided. Theories of the phonetic module emphasize speaker-specific variation (e.g. Kingston and Diehl 1994), and so the need for teasing apart phonetic and phonological contributions to perceived speech sounds has become apparent. There is also greater understanding of cognitive and articulatory individual variation and trauma, and it is far easier to record and machine-analyze speech today than it was in 1965. It will no doubt become even easier in the future (e.g. cheap and easy palatography, ultrasound, and nasal airflow measurement would be very useful), and hopefully techniques will develop that will eliminate much more of the uncertainty that still exists currently. It is therefore important to be frank about our current best capabilities; they are far from certain because no technique provides a clear picture of the PhM at work. It is probably inevitable, then, that even the best of the most recent descriptions will one day be considered inadequate evidence.

So, while EC's description in 1965 was probably as good as a Generative phonologist could have hoped for, the only relevant question now is whether EC's description remains reliable and relevant given current expectations. As the preceding section suggested, with all the developments of the past half-century EC's description now has a very uncertain status as evidence for a Generative theory.

One way to approach development of evidentiary standards is to try to codify the most stringent requirements that reviewers place on journal articles. For example, I suspect there is a consensus that a phonological description based on primary fieldwork should include information about the subjects (e.g. sex, age, perhaps native speaker status), and that there should be at least some phonetic analysis – I suspect that impressionistic transcription (at least for some properties) may nowadays be less frequent in journal articles, or at least reports of phonetic analysis are more common (see Rice 2011's section on "Language Documentation" for relevant references).

The great danger is that requirements will be explicitly or implicitly imposed that are not derived from a theory. For example, while literature on language documentation and fieldwork proposes a number of desiderata for subjects and descriptions, evidence requirements are not often shown to derive from a theory; they are often prescriptive or cater to non-theoretical needs as the goals are often to document languages for social purposes (e.g. Crowley 2007). For example, Rice (2005) and Noonan (2005) identify a number of properties of "good [descriptive] grammars", observing that grammars have many different audiences apart from Generative linguists and roles other than providing evidence for Generative theories. So, their proposed desiderata are not derived from Generative theories, but rather seek to address the conflicting demands of the description's readership. Cheliah and de Reuse (2011:ch.7) present an extensive discussion of the ideal fieldwork subject. Many of their criteria are similar to those required in Generative work. However, it is not clear from what principles their criteria derive; it seems that there is in part a sociological aspect to their proposals as they imply the goal is language documentation (rather than determining possible states of cognitive modules).

The final issue is responsibility for evaluation. At no point was it suggested here that Echeverría and Contreras failed in their task because they did not address particular issues such as the speaker's impairments, and so on. In fact, no part of the preceding discussion was – or should be read as – a criticism of EC's description. It instead advocates the view that evaluating a source is an essential part of using it as evidence for a theory. However, who is responsible for evaluating sources?

The source's author cannot be held responsible. EC did not necessarily have modern Generative theories in mind when they wrote their description, and so it is impossible to fault them for not addressing issues that arise from Generative theories. Responsibility must fall on the theoretician who repurposes EC's description as evidence for their theory. So, if I cite EC's description as evidence for a metrical theory set in OT, it is incumbent on me to show that EC's description is accurate and relevant for its new use as evidence for my theory. Even if the description is set within the same theoretical framework as my theory, it is still necessary to demonstrate that the description is accurate and relevant; adoption of evidence necessarily places responsibility for evidence evaluation on the adopter. In other words, it is not EC's failure that they did not report their subjects' native speaker status, or that they pooled their data; it would be my failure if I used their description for my theory when my theory required subjects to be native speakers and to not pool data.

Perhaps part of the problem relates to perception of the status of descriptions such as EC's. Articles that cite sources without providing a detailed evaluation of them are assuming that the description is not only a fact, but a Generative fact. However, descriptions are merely hypotheses for which (hopefully) adequate and relevance evidence is available. The first questions for any theoretical work that cites a source, then, are whether the source's description can be recast as a hypothesis in that theoretical framework, and whether there is evidence for the hypothesis.

The larger problem touched on here is that recovering information about the output state of a deeply embedded cognitive module from data that has undergone multiple metamorphoses is incredibly difficult (Chomsky 1965: 18-19, 1971: 130). Currently, there is no way to detect phonological inputs or outputs directly, and at least in some cases the procedures that are used have not been proven to be valid. It is to be expected, then, that if any description is presented as evidence for a Generative theory, it should be accompanied by an extensive justification for its use.

A potential concern is that if the bar is set too high for evidence, particularly for stress, there will be no evidence left and so no way to evaluate theories. However, this concern may be unwarranted; without a thorough review of the evidence and concrete proposals about evidentiary standards, there is no way to know which cases meet what standard. In any case, I do not fear discovering that we know nothing with confidence. What I do fear is that significant generalizations about PhMs have been missed and theoretical development has been derailed because inadequate, irrelevant descriptions have been assumed to be valid evidence.

References

Akpinar, Meltem Esen, Ismail Kocak, Berk Gurpinar, and Halil Erturk Esen 2010. Effects of soft palate implants on acoustic characteristics of voice and articulation. *Journal of Voice* 25.3: 381-386.

Alderete, John 2001. *Morphologically governed accent in Optimality Theory* Psychology Press.

Allen, W. Sidney 1975. *Accent and Rhythm*. Cambridge Studies in Linguistics 12. Cambridge: CUP.

Altshuler, Daniel 2009. Quantity insensitive iambs in Osage. *International Journal of American Linguistics* 75.3: 365-398.

Archibald, John 1995. "The acquisition of stress," in John Archibald (ed.) *Phonological acquisition and phonological theory*. Hillsdale, NJ: Lawrence Erlbaum associates, pp. 81-110.

Augusta, Fray Félix José de 1903. Gramática Araucana. Valdivia: Imprenta Central, J. Lampert.

Babor, T, J Higgins-Biddle, J Saunders, and M Monteiro 2001. *AUDIT: The alcohol use disorders identification test* World Health Organization.

Barnes, Jonathan. 2002. *Positional neutralization: A phonologization approach to typological patterns*, Doctoral dissertation, University of California, Berkeley.

Beck, Janet Mackenzie 1997. "Organic variation of the vocal apparatus," in William Hardcastle and John Laver (eds.) *The handbook of phonetic sciences*. Oxford: Blackwell, pp. 256-297.

Beckman, Mary, and Jan Edwards 1990. "Lengthenings and shortenings and the nature of prosodic constituency," in J. Kingston and M. Beckman (eds.) *Papers in Laboratory Phonology I*. Cambridge: Cambridge University Press, pp. 179-200.

Beckman, Jill N. 1998. *Positional faithfulness*, Doctoral dissertation, University of Massachusetts, Amherst.

Berko, Jean 1958. The child's learning of English morphology. *Word* 14: 150-177.

Bernthal, John, Nicholas Bankson, and Peter Flipsen 2009. *Articulation and phonological disorders* Allyn & Bacon.

Best, Catherine 1999. "Development of language-specific influences on speech perception and production in pre-verbal infancy," in *Proceedings of the XIVth International Congress of Phonetic Sciences*. San Francisco, CA., pp. 1261-1264.

Bley-Vroman, R. 1988. "The fundamental character of foreign language learning," in W. Rutherford and M. Sharwood-Smith (eds.) *Grammar and Second Language Teaching: a Book of Readings*. New York: Newbury House/Harper & Row, pp. 19-30.

Boothroyd, Arthur 1984. Auditory perception of speech contrasts by subjects with sensorineural hearing loss. *Journal of Speech and Hearing Research* 27: 133-144.

Broselow, Ellen, Su-I Chen, and Marie Huffman 1997. Syllable Weight: Convergence of Phonology and Phonetics. *Phonology* 14.1: 47-82.

Broselow, Ellen 2009. "Stress adaptation in loanword phonology: Perception and learnability," in Paul Boersma and Silke Hamann (eds.). Berlin: Mouton de Gruyter, pp. 191-234.

Browman, C. P., and L. M. Goldstein 1995. "Dynamics and articulatory phonology," in R.F. Port and T. van Gelder (eds.) *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press, pp. 175-193.

Burzio, Luigi 1994. *Principles of English stress*. Cambridge Studies in Linguistics 72. Cambridge: CUP.

Butterworth, Brian 1979. Hesitation and the production of verbal paraphasias and neologisms in jargon aphasia. *Brain and Language* 8: 133-161.

Bye, Patrik, and Paul de Lacy 2008. "Metrical influences on lenition and fortition," in Joaquim de Carvalho, Tobias Scheer and Philippe Ségéral (eds.) *Lenition and Fortition*. Berlin: Mouton de Gruyter, pp. 173-206.

CDC (US Centers for Disease Control) 2008. *Tooth decay, untreated: US, 1988-2008 (Source: NHANES)*. http://205.207.175.93/HDI/TableViewer/tableView.aspx?ReportId=106 [Accessed:30 March, 2012].

Cheliah, Sobhana and Willem de Reuse 2011. *Handbook of descriptive linguistic fieldwork*. New York: Springer.

Chen, Matthew 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22: 129-159.

Chin, Steven, and David Pisoni 1997. *Alcohol and speech*. San Diego: Academic Press.

Cho, Taehong, and Patricia Keating 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37.4: 466-485.

Choi, John 1995. An acoustic-phonetic underspecification account of Marshallese vowel allophony. *Journal of Phonetics* 23.3: 323-347.

Chomsky, Noam, and Morris Halle 1968. *The sound pattern of English*. New York: Harper & Row.
Chomsky, Noam 1971. "Formal discussion: the development of grammar in child language," in J.P.B. Allen and P. van Buren (eds.) *Chomsky: Selected Readings*. Oxford: Oxford University Press, pp. 129-134.
Coetzee, Andries, and Joe Pater 2011. "The place of variation in phonological theory," in John Goldsmith, Jason Riggle and Alan Yu (eds.) *The handbook of phonological theory*. Oxford: Blackwell, pp. 401-434.
Cohn, Abigail 2011. "Laboratory Phonology: Past successes and current questions, challenges, and goals," in C. Fougeron, B. Kühnert, M. D'Imperio and N. Vallé (eds.) *Laboratory Phonology 10*. Berlin: Mouton De Gruyter, pp. 3-29.
Cook, Vivian 2000. "Linguistics and second language acquisition: One person with two languages," in M. Aronoff and J. Rees-Miller (eds.) *The Blackwell handbook of linguistics*: Wiley-Blackwell, pp. 488-511.
Crosswhite, Katherine 2004. "Vowel reduction," in Bruce Hayes and Donca Steriade (eds.) *Phonetically based phonology*. Cambridge: Cambridge University Press, pp. 191-231.
Crowley, Terry 2007. *Field linguistics: a beginner's guide*. Oxford: Oxford University Press.
Cummings, Louise 2008. *Clinical Linguistics*. Edinburgh: Edinburgh University Press.
Davies, William, and Tamar Kaplan 1998. Native speaker vs. L2 learner grammaticality judgements. *Applied Linguistics* 19.2: 183-203.
de Jong, Kenneth 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 97.1: 491-504.
de Lacy, Paul 2004. Markedness conflation in Optimality Theory. *Phonology* 21.2: 145-199.
de Lacy, Paul 2007. The interaction of tone, sonority, and prosodic structure. In Paul de Lacy (ed.) *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press, pp. 281-307.
de Lacy, Paul 2009. "Phonological evidence," in Steve Parker (ed.) *Phonological argumentation: Essays on evidence and motivation*: Equinox Publications, pp. 43-77.
den Ouden, Bart. 2002. *Phonology in Aphasia*, Doctoral dissertation, Rijksuniversiteit Groningen.
Echeverría, Sergio (Max) 1964. Descripción fonológia del mapuche actual. *Boletín del Instituto de Filología de la Universidad de Chile XVI*: 13-59.
Echeverría, Max S., and Heles Contreras 1965. Araucanian Phonemics. *International Journal of American Linguistics* 31.2: 132-135.
Fabre, Alain 1998. *Manual de las lenguas indígenas sudamericanas*. Munich: Lincon Europa.
Genesee, Fred, and Elena Nicoladis 2007. "Bilingual first language acquisition," in Erika Hoff and Marilyn Shatz (eds.) *Handbook of Language Development*. Malden, MA: Blackwell Publishing, pp. 324-342.
Gordon, Matthew 1999. *Stress and Other Weight-Sensitive Phenomena: Phonetics, Phonology, and Typology* Doctoral Dissertation, UCLA.
Gordon, Matthew 2000. Re-examining default-to-opposite stress. *Proceedings of the Annual Meeting of the Berkeley Linguistic Society* 26:101-112. http://elanguage.net/journals/bls/issue/view/361
Gordon, Matthew 2002. A factorial typology of quantity-insensitive stress. *Natural Language and Linguistic Theory* 20: 491-552.

Gordon, Matthew 2004. A phonological and phonetic study of word-level stress in Chickasaw. *International Journal of American Linguistics* 70: 1-32.
Gordon, Matthew 2011. "Stress systems," in J. Goldsmith, A. Yu and J. Riggle (eds.) *The Handbook of Phonological Theory*. Second Edition. Oxford: Blackwell, pp. 141-163.
Gu, Wentao, Tan Lee, and P.C. Ching 2008. Prosodic variation in Cantonese-English code-mixed speech. *International symposium on Chinese spoken language processing (ISCSLP 2008)*: 342-345.
Gussenhoven, Carlos 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
Gussenhoven, Carlos 2007. "The phonology of intonation," in Paul de Lacy (ed.) *The Cambridge Handbook of Phonology*. Cambridge: CUP, pp. 253-280.
Harris, Zellig S. 1960. *Structural linguistics*. Chicago: University of Chicago Press.
Hartshorne, Joshua, and Adena Schachner 2012. Tracking replicability as a method of post-publication open evaluation. *Frontiers in Computational Neuroscience* 6.8.
Hayes, Bruce. 1981. *A metrical theory of stress rules*, Doctoral dissertation, MIT.
Hayes, Bruce 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: The University of Chicago Press.
Haynes, William, and Rebekah Pindzola 2011. *Diagnosis and evaluation in speech pathology* Allyn & Bacon.
Heinz, Jeffrey. 2012. UD Phonology Lab Stress Pattern Database. Version 03/22/2012. http://phonology.cogsci.udel.edu/dbs/stress/
Hung, Henrietta. 1994. *The rhythmic and prosodic organization of edge constituents*, Doctoral dissertation, Brandeis University.
Hyde, Brett. 2001. *Metrical and prosodic structure in Optimality Theory*, Doctoral dissertation, Rutgers University. Rutgers Optimality Archive 476.
Hyde, Brett 2002. A Restrictive Theory of Metrical Stress. *Phonology* 19: 313-339.
Hyman, Larry 1977. "On the nature of linguistic stress," in Larry Hyman (ed.) *Studies in stress and accent*, pp. 37-82.
Idsardi, William, and David Poeppel 2010. "Neurophysiological and neuroimaging techniques: general advantages and disadvantages," in Cecile Fougeron Abigail Cohn, and Marie Huffman (ed.) *Handbook of Laboratory Phonology*. Oxford: Oxford University Press, pp. 593-605.
Jastreboff, Pawel 1990. Phantom auditory perception (tinnitus): mechanisms of generation and perception. *Neuroscience Research* 8.4: 221-254.
Kager, René 1993. Alternatives to the iambic-trochaic law. *Natural Language and Linguistic Theory* 11: 381-432.
Kager, René 2007. "Feet and metrical stress," in Paul de Lacy (ed.) *The Cambridge Handbook of Phonology*. New York: Cambridge University Press, pp. 195-228.
Kawahara, Shigeto 2011. Japanese loanword devoicing revisited: a rating study. *Natural Language and Linguistic Theory* 29.3: 705-723.
Kawahara, Shigeto 2011. "Experimental approaches in theoretical phonology," in M. van Oostendorp, C. Ewen, E. Hume and K. Rice (eds.) *The Blackwell Companion to Phonology*: Wiley-Blackwell, pp. 2283-2303.
Keating, Patricia 1990. Phonetic representations in a generative grammar. *Journal of Phonetics* 18: 321-334.

Ketner, Katherine 2006. Size restrictions in prosodic morphology. Doctoral dissertation, University of Cambridge. Rutgers Optimality Archive 1028.

Kingston, John, and Randy Diehl 1994. Phonetic knowledge. *Language* 70: 419-454.

Kingston, John 2007. "The Phonetics-Phonology Interface," in Paul de Lacy (ed.) *The Cambridge Handbook of Phonology*. Cambridge: CUP, pp. 401-436.

Kiparsky, Paul 1982. "Lexical phonology and morphology," in I. S. Yang (ed.) *Linguistics in the Morning Calm*. Seoul: Hanshin, pp. 3-91.

Kirchner, Robert 2001. *An effort based approach to consonant lenition* Psychology Press.

Labov, William 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William 1984. "Field methods of the project on linguistic change and variation," in J. Baugh and J. Sherzer (eds.) *Language in use*. Englewood Cliffs: Prentice Hall, pp. 28-53.

Labov, William 2006. *The social stratification of English in New York City*. Cambridge: Cambridge University Press.

Ladefoged, Peter 2003. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Malden, MA: Blackwell Publishing.

Lea, W. A. 1973. "Segmental and suprasegmental influences on fundamental frequency contours," in Larry Hyman (ed.) *Consonant Types and Tones*. Los Angeles: University of Southern California Press, pp. 15–70.

Lehiste, Ilse 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.

Lindblom, Björn 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35.11: 1773-1781.

Löfquist, Anders 1997. "Theories and models of speech production," in William Hardcastle and John Laver (eds.) *The handbook of phonetic sciences*. Oxford: Blackwell, pp. 405-426.

Maddieson, Ian, and Kristen Precoda 1992. *UPSID and PHONEME (user guide).* UCLA Linguistics Department. http://www.linguistics.ucla.edu/faciliti/sales/software.htm.

McCarthy, John, and Alan Prince 1993. "Generalized Alignment," in Geert Booij and Jaap van Marle (eds.) *Yearbook of Morphology*. Dordrecht: Kluwer, pp. 79-153.

McGarrity, Laura. 2003. *Constraints on patterns of primary and secondary stress*, Doctoral dissertation, Indiana University.

Milroy, Leslie 1987. *Observing and analysing natural language*. Oxford: Basil Blackwell.

Milroy, Leslie, and Matthew Gordon 2003. *Sociolinguistics: Method and Interpretation*. Malden, MA: Blackwell Publishing.

Molineaux, Benjamin . 2012. The prosodic word in Mapudungun: a stress-based account. Poster presented at the Fifth Tone and Intonation in Europe Conference (TIE 5). 7 September 2012.

Mosel, Ulrike, and Even Hovdhaugen 1992. *Samoan reference grammar*. London, Tromso: Oxford University Press.

Nace, Edgar 2005. "Alcohol," in Richard J. Frances, Sheldon I. Miller and Avram H. Mack (eds.) *Clinical textbook of addictive disorders*. NY & London: Guildford Press, pp. 75-104.

Noonan, Michael 2005. Grammar writing for a grammar-reading audience. *Studies in Language* 30.2: 351-365.

Ohala, John 1978. "Production of tone," in V. Fromkin (ed.) *Tone: A linguistic survey*. New York: Academic Press, pp. 5-39.

Ohala, John 1986. Consumer's guide to evidence in phonology. *Phonology Yearbook* 3: 3-26.

O'Keefe, Michael 2007. *Phonetic correlates of stress in Mapudungun*: Handout. Rutgers University. http://www.pauldelacy.net/web/docs/okeefe.pdf.

Pankratz, Leo, and Eunice V. Pike 1967. Phonology and morphtonemics of Ayutla Mixtec. *International Journal of American Linguistics* 33: 287-299.

Parker, Steve 1999. A sketch of Iñapari phonology. *International Journal of American Linguistics* 65.1: 1-39.

Peperkamp, Sharon, and Emmanuel Dupoux 2002. "A typological study of stress 'deafness'," in Carlos Gussenhoven and Natasha Warner (eds.) *Laboratory Phonology 7*. Berlin: Mouton de Gruyter, pp. 203-240.

Pinchoura-Fuller, M. Kathleen, and Pamela E. Souza 2003. Effects of aging on auditory processing of speech. *International Journal of Audiology* 42: 2S11-12S16.

Prince, Alan 1983. Relating to the grid. *Linguistic Inquiry* 14: 19-100.

Prince, Alan, and Paul Smolensky 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Wiley-Blackwell.

Quick, Phil 2004. "Creaky voice as a phonetic manifestation of the glottal stop in Pendau," in Andrea Rakowski and Norvin Richards (eds.) *Proceedings of AFLA VIII: the Eighth Meeting of the Austronesian Formal Linguistics Association*. Cambridge, MA: MIT Linguistics Department, pp.

Rice, Keren 2005. A typology of good grammars. *Studies in Language* 30.2: 385-415.

Rice, Keren 2011. "Fieldwork," in *Oxford Bibliographies Online: Linguistics*, pp.

Rohrer, Jonathan, Martin Rossor, and Jason Warren 2009. Neologistic jargon aphasia and agraphia in primary progressive aphasia. *Journal of Neurological Science* 277.1-2: 155-159.

Salamanca, Gastón and Elizabeth Quintrileo 2009. El mapuche hablado en Tirúa: Fonemas segmentales, fonotaxis y comparación con otras variedades. *RLA, Revista de Lingüística Teórica y Aplicada* 47.1: 13-35.

Salas, Adalberto 1992. *El Mapuche o Araucano: Fonología, gramática y antología de cuentos*. Madrid: Editorial MAPFRE.

Sapir, Edward 1930. Southern Paiute, a Shoshonean language. *Proceedings of the American Academy of Arts and Sciences* 65: 1-296.

Selzer, M.L. 1971. The Michigan Alcoholism Screening Test; the quest for a new diagnostic instrument. *The American Journal of Psychiatry* 127: 1653-1658.

Smeets, Ineke 1989. *A Mapuche Grammar*, Doctoral dissertation, Universiteit Leiden.

Stemberger, Joseph, and Barbara Bernhardt 2007. "Phonological impairment in children and adults," in Paul de Lacy (ed.) *The handbook of phonology*. Cambridge: CUP, pp. 575-593.

Sutton, Jon (ed.) 2012. Replication, replication, replication. *The Psychologist* 25.5.

Tesar, Bruce 2004. Using inconsistency detection to overcome structural ambiguity in language learning. *Linguistic Inquiry* 35.2: 219-253.

Truckenbrodt, Hubert 2007. "The syntax-phonology interface," in Paul de Lacy (ed.) *The Cambridge Handbook of Phonology*. Cambridge: CUP, pp. 435-456.

Trudgill, Peter 1999. "Standard English: What it isn't," in Tony Bex and Richard J. Watts (eds.) *Standard English: the widening debate*. London: Routledge, pp. 117-128.

UNPO (Underrepresented Nations and Peoples Organization) 2008. *Mapuche*. http://www.unpo.org/members/7895 [Accessed:30 March, 2012].

Ussishkin, Adam 2007. "Morpheme position," in Paul de Lacy (ed.) *The Cambridge Handbook of Phonology*. Cambridge: CUP, pp. 457-472.

Weismer, Janet Mackenzie 1997. "Motor speech disorders," in William Hardcastle and John Laver (eds.) *The handbook of phonetic sciences*. Oxford: Blackwell, pp. 191-219.

Welfare, Australian Institute of Health and 2007. *Statistics on drug use in Australia 2006*. Canberra: Australian Institute of Health and Welfare.

Werker, Janet, and Richard Tees 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7.1: 49-63.

Whalen, D.H., and A.E. Levitt 1995. The universality of intrinsic f_0 of vowels. *Journal of Phonetics* 23: 349-366.

Williams, Herbert W. 1917. *A Dictionary of the Maori Language*. New Zealand: Government Printer.

Wolfram, Walt 2011. "Fieldwork methods in language variation," in Ruth Wodak, Barbara Johnstone and Paul E. Kerswill (eds.) *The SAGE Handbook of sociolinguistics*. London: SAGE Publications, pp. 296-312.

Zeng, Fan-Gang, and Christopher Turner 1990. Recognition of voiceless fricatives by normal and hearing-impaired subjects. *Journal of Speech and Hearing Research* 33: 440-449.

Ziegler, Wolfram, and Ben Maassen 2004. "The role of the syllable in disorders of spoken language production," in Raymond Kent Ben Maassen, Herman Peters, Pascal van Lieshout and and Wouter Hulstijn (eds.) *Speech motor control in normal and disordered speech*. Oxford: Oxford University Press, pp. 415-447.

Zúñiga, Fernando 2000. *Mapudungun*. Munich: Lincom Europa.

Zúñiga, Fernando 2006. *Mapudungun: El habla mapuche*. Santiago de Chile: Centro de Estudios Públicos.