Phonetic imitation, accommodation, and audience design in Canadian radio interviews[1]

Jennifer Abel
Department of Linguistics, University of British Columbia

This study examined the phenomenon of spontaneous phonetic imitation in one Canadian English talker's speech through a corpus of podcasted radio interviews. No evidence was found of imitation in either the two variables which were examined: the low back vowels involved in the *cot-caught* merger or the Canadian Raising diphthong [aɪ]. While imitation was not found, it is possible that sociolinguistic adaptation such as accommodation (Giles and Ogay 2007) or audience design (Bell 1984) could have contributed to the changes in the talker's productions over time, depending on whether the talker was adapting to his immediate interlocutor or to his broader listening audience. Ultimately, uncontrolled conversations may not be fruitful places to look for true imitative behaviour of the type described by Goldinger (1998), *inter alia*; nevertheless, previously untapped data sources such as podcasts may be important in future research into sociolinguistic and sociophonetic phenomena.

1 Introduction

In any interpersonal spoken interaction, a number of factors at a variety of linguistic and social levels will influence the phonetic characteristics of a talker's speech. At a fairly basic social level, *spontaneous phonetic imitation* involves the speech of some talker A becoming more similar to that of another talker B following A's exposure to B's speech, without conscious effort or intention by A. A range of studies have suggested that, at least in controlled laboratory settings, spontaneous phonetic imitation is a robust phenomenon in speech behaviour, affecting such things as fundamental frequency (Gregory 1990; Gregory et al. 1993), vocal intensity (Natale 1975), word duration (Abel et al. 2011), vowel quality (Babel 2010, 2012), and VOT (Abrego-Collier et al. 2011; Nielsen 2011; Shockley et al. 2004). In turn, imitation is influenced by such factors as word frequency (Goldinger 1998), level of attention to the stimulus (Abel et al. 2011), gender (Namy et al. 2002; Pardo 2006), attitudes towards the model talker (Abrego-Collier et al. 2011; Babel 2010, 2012), social conformity (Natale 1975), conversational role (Pardo 2006), and inter-speaker dialect distance (Kim et al. 2011).

Imitation research often overlaps with that on speech *accommodation*: socially-motivated adaptation of vocal behaviour to create, maintain, or reduce distance between interlocutors or identification with a particular group (Giles and Ogay 2007). Accommodation research explores both convergent and divergent behaviour, rather than imitation per se, generally in one-on-one interactions (e.g., Bilous and Krauss 1988; Coupland 1980; Giles 1973; Giles et al. 1973; Gregory and Webster 1996). When larger groups of addressees – both direct and indirect – are involved (e.g., Gregory and Gallagher 2002; Hall-Lew et al. 2010; Hay et al. 1999), elements of *audience design* or *referee design* (Bell 1984) may also come into play, as talkers shift their speech behaviour to not only take account of their immediate interlocutors, but also various kinds of non-participatory third parties (present third parties are an audience, non-present third parties are referees). Accommodation and audience design can take place before an interaction begins (i.e., a talker predicting what an interlocutor's or audience member's speech behaviour will be like and adjusting based on that prediction), as well as taking place during the course of an interaction (Giles 1973); imitation, on the other hand, only takes place after exposure (i.e., during or after an interaction). As with imitation, talkers have been found to accommodate or incorporate audience/referee design in phonetic aspects of speech like fundamental frequency (Gregory and Webster 1996; Gregory and Gallagher 2002), stop voicing or flapping (Bell 1984; Coupland 1980) and vowel

[1] Many thanks to my committee members, Carla Hudson Kam, Molly Babel, and Lisa Matthewson, and to Bryan Gick and my classmates in the Production and Perception class in 2011-2012.

quality (Coupland 1980, 2001; Hall-Lew et al. 2010; Hay et al. 1999), as well as at a variety of other linguistic levels. Likewise, accommodation has been found to be affected by factors like gender (Bilous and Krauss 1988) and word frequency (Hay et al. 1999), as well as broader social factors like dominance (Gregory and Webster 1996), race (Hay et al. 1999), and political affiliation (Hall-Lew et al. 2010).

In controlled settings such as laboratory experiments or sociolinguistic interviews, it is possible to remove some of the possible interaction between factors like these to some degree. In imitation studies, subjects often listen to recordings of single words by an unknown talker multiple times and then repeat those single words. Alternately, they may participate in task-specific dyadic interactions (e.g., find differences between pictures; draw a route on a map) with limited social interaction between the interlocutors (conversations are limited to on-task topics, measures are taken to minimize visual or physical interaction between the talkers). In interviews, talkers often converse with a single individual on a single occasion about some range of topics designed by the interviewer to examine the talker's language use; as well, the observer's paradox (that the interviewer will inevitably have an effect on the talker's speech) is almost always in play. But outside of these settings, any or all of the factors mentioned above – plus a wide range of others – could be behind variation in a talker's speech, and the interaction between them could create patterns which are not within the scope of the predictions made in controlled environments. In particular, spontaneous phonetic imitation seems particularly liable to disappear outside the laboratory, given the fairly rigorous controls usually surrounding its documentation. All of this is further complicated by the difficulty of observing these interactions 'in the wild', as it were, given the difficulty of recording 'natural' conversations in a way which circumvents the observer's paradox.

This study examines sociophonetic variation in one Canadian English talker's productions from two vowel categories within and across a number of live-to-tape radio interviews to attempt to determine whether this variation is due to spontaneous phonetic imitation. In contrast to laboratory research, which has found imitation when subjects are exposed to multiple repetitions of lexical items in isolation or engage in limited-vocabulary conversations, this study uses uncontrolled conversations, with words which may or may not have multiple repetitions, with variable amounts of time between talkers' conversational turns, and in a broader discourse context with no *a priori* limits on the vocabulary choices. Because of the nature of these conversations, and with the aim of looking at natural speech interactions, some of the factors which have been found to influence imitation in the laboratory – in particular, control over word frequency (low-frequency words are imitated more than high-frequency words (Goldinger 1998)) and latency between exposure and production (a great deal of imitation research has used an immediate shadowing paradigm, where subjects repeat words immediately after hearing the model talker pronounce them, although imitation has also been found after delays between exposure and repetition (e.g. Abel et al. 2011; Pardo 2006) – were not controlled for. Whether imitation would occur in the absence of control of these factors is one of the questions under investigation. As well, the rich conversational environment introduces many more possible causes of variation than are present in the laboratory setting, both social and linguistic. In particular, the possible impact of accommodation and audience design on creating convergent behaviour will be considered in conducting the analysis. If spontaneous phonetic imitation is not simply an artifact of exposure to speech stimuli in controlled laboratory conditions, we would expect to find it in at least some naturally-occurring conversational situations with at least some speech variables. This study represents a very early step in exploring what those situations and variables could be.

The conversations were taken from a corpus of interviews collected from podcasts of the CBC Radio cultural affairs program *Q*. An additional goal of this study is to demonstrate the potential usefulness of podcasts in linguistic research generally and phonetic research particularly; using the CBC's podcast archive allowed for the compilation of a sizable corpus of downloadable (and thus easily manipulable) broadcast-quality recordings at no cost. Corpora compiled from broadcast material (although not from podcasts) have been used previously in both accommodation and audience/referee design research. Most of the corpus work has focused on talkers modulating some aspect of their speech in response to multiple addressees, audience members, or referees, including Bell's (1984) work on newscaster speech in New Zealand, Gregory and Gallagher's (2002) examination of candidates' f0 changes in U.S. election debates, Hay et al.'s (1999) work on [aɪ] monophthongization in Oprah Winfrey's introductions of guests, and Hall-Lew et al.'s (2010) research on U.S. congresspeople's pronunciation of the second vowel in <Iraq> during debates. The latter two studies, which both examine vowel quality, only make categorical distinctions between pronunciations (i.e., monophthongized or not, [æ] or [ɑ]) rather than gradient ones, and do not look at changes over time. The only corpus study which looks at conversational interview material seems to be Gregory and Webster (1996), which examines convergence and divergence in fundamental frequency by Larry King and his guests. Broadcast corpora do not appear to have been used in phonetic imitation research; whether they will prove suitable for investigations this area is another of the questions addressed by this study.

Q is hosted by Jian Ghomeshi (hereafter, JG), a London, England-born Canadian of Iranian descent (Canadian Broadcasting Corporation 2011). JG immigrated to the Greater Toronto Area around the age of 7 (Brown 2003), and has lived in that area since then. Interviewees included both men and women speaking a variety of dialects of English. While some controls were used to determine what material would be analyzed (see sections 2.1 and 2.2 for details), these controls were much less stringent than those used in previous spontaneous phonetic imitation studies. Two variables were selected for analysis in this preliminary study, both of which might be expected to be different in JG's dialect than in those of (at least) some of his interviewees (Boberg 2008; Labov et al. 2006; Wells 1982): the Canadian Raising diphthongs [aɪ] and [aʊ] (Chambers 1989, 2006; Joos 1942), and the vowels involved in the *cot-caught* merger (Labov et al. 2006; Wells 1982). If JG is imitating his interviewees' productions of these vowels, we would expect that within a given interview, JG's productions should become more like those of his guest as the interview progresses; e.g., if the guest is a speaker of a dialect without Canadian Raising, JG should display less Canadian Raising at the end of the interview than he does at the beginning of the interview because of his exposure to the guest's non-raised productions.[2] If JG is accommodating to his interviewees, on the other hand, it is possible that his productions may converge to theirs over time, but it is also possible that they may diverge or even remain the same, depending on whether JG is trying to reduce, increase, or maintain social distance. To a larger degree than imitation, accommodative behaviour is likely to be driven by factors such as nationality or relative socioeconomic status of the interlocutors. For this reason, analysis was planned for within interviews, within merger and raising groups, and within national groups. As well, if the needs or expectations of the listening audience prevail in terms of JG's behaviour, he could again display convergence, divergence, or maintenance in his productions over time. Thus, it may not always be possible to determine whether increasing similarity in productions is due to imitative behaviour alone; however, if divergence is found between JG's productions and those of his interlocutors, it can reasonably be concluded that imitation is not occurring.

2 Methods

2.1 Interview/subject selection

Podcasts were downloaded from http://www.cbc.ca/podcasting/includes/qpodcast.xml. To control beforehand for at least some of the variability inherent in using uncontrolled material, the following criteria were used in selecting interviews. The interviewer had to be JG; this eliminated many of the interviews which aired on July or August programs, as guest hosts are typically used during those months. To control for some of the possible factors which might affect the course or naturalness of the conversations – such as transmission issues (e.g., poor quality audio, delays in responses, loss of telephone lines), editing (e.g., cutting or rearranging the material in an interview for reasons of content/time), distractions affecting one interlocutor and not the other, *inter alia* – the interviewee had to be live in the studio on the broadcast day. Thus, no pre-recorded interviews or previously broadcast interviews were selected, nor were telephone or studio-to-studio interviews. To control for effects of interactions with multiple interlocutors, no other people could be involved in the interview besides JG and the interviewee; this eliminated panel discussions and multi-interviewee interviews. The interviewee had to be a native speaker of English; this was assessed both through auditory information from the interview and through collecting biographical information about the interviewee (from the interview itself, or from

[2] This will, of course, depend to some degree on what JG's 'normal' productions are; if he is not normally a significant Canadian Raiser, he will not necessarily display a dramatic change in his productions when speaking with a non-Raiser, but some change might be expected.

publicly-available sources such as Wikipedia or the Internet Movie Database (IMDB)). Using these criteria, interviews with 12 interlocutors were selected for analysis; to create balance among national groups and to encompass a range of English dialects, the interviewees included four speakers of dialects of Canadian English, four of dialects of American English, and four of dialects of British English. These interviews originally aired on *Q* broadcasts between June 24, 2010 and September 9, 2011. Air dates of the interviews and dialect information for the interviewees is given in Table 1.

Table 1. Dialect information and interview air dates for JG's interlocutors (AM = American dialect grouping (Labov et al. 2006), CA = Canadian dialect grouping (Boberg 2008), UK = British dialect grouping (Wells 1982)).

Guest	**Sex**	**Interview Air Date**	**Interview Length**	**Born/ Formative Years**	**Dialect**
AM	F	29 Sep 2010	18m27s	Portland, ME	AM: Eastern New England
BC	M	13 Oct 2010	22m56s	Glasgow, UK	UK: Scottish
BM	M	24 Jun 2010	16m14s	Kingston, ON	CA: Ontario
CL	F	7 Jul 2010	15m21s	New York, NY	AM: New York City
HW	F	13 Oct 2010	11m14s	Shaunavon, SK	CA: Prairies
IM	M	9 Jul 2011	16m12s	Portland, OR	AM: The West
KF	M	12 May 2011	18m21s	Chicago, IL/Texas	AM: The South (Texas)
ML	M	15 Sep 2010	18m18s	Manchester, UK	UK: Northern (Manchester)
NL	F	2 Dec 2010	21m18s	London, UK	UK: Southern Standard
PC	M	12 Oct 2010	20m50s	London, UK	UK: London working-class
SM	M	9 Sep 2010	15m46s	Burlington, NF	CA: Newfoundland
SP	F	9 Sep 2011	21m14s	Toronto, ON	CA: Ontario

The mean length of the interviews was 18 minutes (SD = 3m17s). Interviews were extracted from the MP3 podcasts using Wavesurfer (Sjölander and Bestow 2006) and saved as WAV files with a sampling rate of 44.1kHz. Inspection of the spectrograms suggests that the podcasts were digitized at 32kHz.

2.2 Lexical item identification, selection and labeling

In selecting the lexical items to be used in the analysis, certain types of material were excluded in order to ensure both maximum clarity of the acoustic signal[3] and maximum conversational naturalness of the speech. Excluded material included read material which was directed towards the listening audience rather than the interviewee (e.g., introductory essays, introductions of guests); utterances which were laughed through, sung, or uttered in a 'character' voice; utterances which overlapped with an utterance/laughter/noticeable breathing by the other interlocutor, or which overlapped with other noises; and clear disfluencies/hesitations/mispronunciations.

Word and vowel tokens were identified using waveform, spectrographic and auditory information, and hand-labeled using Praat (Boersma and Weenink 2009). Canadian Raising tokens were categorized as follows: tokens of [aɪ] or [aʊ] diphthongs in possible Canadian Raising environments (i.e., before tautosyllabic voiceless obstruents and voiceless obstruents in the onset of a following unstressed syllable; see e.g., Boberg 2008; Chambers 1989, 2006; Joos 1942) were labeled as 'CRaj' and 'CRaw', and those [aɪ]/[aʊ] tokens in all environments which should not condition CR (including in word-final open syllables) were labeled as 'notCRaj' and 'notCRaw'. Inclusion of tokens in both CR and non-CR environments was necessary to determine if a talker actually displayed CR behaviour. The first person singular pronoun *I* and its cliticized forms were not included among the tokens, as these productions were extremely difficult to separate from the surrounding utterances.

The categorization for vowel tokens involved in the *cot/caught* merger followed the descriptions of the LOT and THOUGHT lexical sets in Wells (1982). For talkers with fully merged dialects, it was not

[3] This was desirable not only for purposes of acoustic analysis but also in the event that the tokens would later be used for a perceptual task; see Kim et al. (2011: 132) for similar types of material excluded from their stimulus selections.

possible to distinguish these vowels acoustically or auditorily; thus, the representative words listed in Wells (1982) and the orthographic conventions used for writing these vowel sounds were used as guidelines. The low back vowels in those words which shared characteristics with the words listed by Wells (1982: 131) as being part of the LOT lexical set (words with the vowel [ɑ] in General American and [ɒ] in Received Pronunciation) were labeled as 'cot'. The lexical set of the LOT vowel (often called 'short o') includes words with orthographic <o> before many voiced and voiceless obstruents, as well as before [m], [n] and [l]; some words with orthographic <a>, such as *swan*, *quality*, *yacht*, and *watch*; and some words with orthographic <ow>, including *knowledge*. Most instances of orthographic <o> before voiceless fricatives were not included, as these are thought to be part of the CLOTH lexical set (Wells 1982: 136); the exceptions were *possible* (and its derivatives) and *profit*. Orthographic <o> before [ŋ] also belongs to the CLOTH lexical set and was thus not included. The low back vowels in those words which shared characteristics with the words listed by Wells (1982: 145) as being part of the core THOUGHT lexical set (words with the vowel [ɔ] in both Received Pronunciation and General American) were labeled as 'caught'. This included most words with orthographic <au>, <ou>, and <aw> (not *fault*, which is variable in different UK dialect regions); <-all> and <-alk> words; and *bald*, *water*, and *broad*. The preposition *on* was not included in either category; Wells (1982: 473) notes that it has a LOT vowel in the Northern United States and a THOUGHT vowel in the Southern United States, and Labov et al. (2006) suggest that *on* has shifted from having a LOT vowel to having a THOUGHT vowel, but do not indicate whether this shift has occurred in all regions of the U.S. or only some of them.

All tokens which fit into these categories and which were separable from the surrounding utterances were labeled; this means that vowel tokens were included from all word positions (initial, medial, final), and word tokens were included from all utterance positions (initial, medial, final). Following Pardo (2006), word and vowel tokens were classified as 'early' or 'late' based on whether they occurred before or after the temporal midpoint of the interview.

2.3 Formant measurement

F1 and F2 values were extracted using a Praat script at seven points in the course of the vowel: the 5%, 10%, 25%, 50%, 75%, 90%, and 95% marks. Tokens for which the extracted formant values seemed unusual were re-measured by hand and corrected where necessary; those for which accurate and reliable measurements could not be obtained either automatically or by hand were discarded. Tokens which did not have the vowel quality expected in the citation form of a lexical item – e.g., a pronunciation of <thought> as [θaɪt] – were included in the final tally of tokens if the measurements were reliable and the token had not been corrected by the talker as a mispronunciation, as a fair amount of variability is expected in reasonably fast, conversational speech.

The formant values for the *cot* and *caught* vowels were taken from the measurements at the 50% mark. To capture the dynamic nature of the formants in the [aɪ] and [aʊ] diphthongs, the measurement technique described by Moreton and Thomas (2007) was used. Measurements were taken of F1 and F2 at both the point of maximum F1 (typically found in the nucleus of the diphthong) and the point of maximum F2 (typically found in the offglide). Across JG's tokens (and those of his interlocutors), the maximum F1 was found at either the 25% mark or the 50% mark of the vowel, and the maximum F2 was found at either the 90% mark or the 95% mark. Thus, the F1 and F2 values for [aɪ] and [aʊ] were taken from the measurements at the 25%, 50%, 90%, and 95% points.

3 Analysis and results

3.1 *Cot/caught* merger

The 12 interviewees were grouped according to the degree of merger of their *cot* and *caught* vowels; this was assessed by measuring the difference between their mean formant values for each vowel type at the 50% point of the vowels. The resulting three groupings and the mean F1 and F2 differences for those groups are given in Table 2.

Table 2. Groupings of interviewees by degree of merger of *cot* and *caught* vowels.

Group	Guests	Dialect Groups	Mean F1 Difference (Hz)	Mean F2 Difference (Hz)
Merged	AM, HW, SP	AM, CA	12.78	22.95
Unmerged	CL, ML, NL, PC	AM, UK	174.85	249.6
Somewhat Merged	BC, BM, IM, KF, SM	AM, CA, UK	42.7	143.44

Welch's two-sample t-tests on the F1 and F2 means for *cot* and *caught* showed that the means for the guests in the Unmerged and Somewhat Merged groups were significantly different, but those for the guests in the Merged group were not. An examination of the distributions for Unmerged and Somewhat Merged guests showed a greater degree of overlap between *cot* and *caught* formant values for the Somewhat Merged guests than for the Unmerged guests. Labov et al. (2006) also found a three-way merger distinction in both production (as judged by the analyst) and perception (as reported by the talker) of the *cot* and *caught* vowels among their subjects, which they labeled as 'same' or 'merged' (=Merged), 'different' or 'distinct' (=Unmerged), and 'close' or 'transitional' (=Somewhat Merged).

The total number of tokens of *cot* and *caught* vowels for JG is given in Table 3.

Table 3. Number of tokens of *caught* and *cot* vowels for JG across all interviews.

Number of tokens	caught	cot
Early	42	115
Late	57	97
Total	99	212
Average per interview	8.25	17.67
Maximum in one interview	16	41
Minimum in one interview	4	8

In general, the number of tokens for JG per interview was not particularly high; in comparison, the average number of *cot* vowel tokens per guest per interview was 46.33 (25.5 early, 24.1 late), and the average number of *caught* vowel tokens was 25.08 (13.4 early, 11.7 late). The number of *caught* tokens for JG in particular was lower than desirable for conducting within-interview analyses.[4] It was thus deemed infeasible to examine vowel quality changes within the individual interviews. Instead, JG's vowel qualities were measured in three larger groupings: across all interviews, across the three merger groupings (Merged, Unmerged, and Somewhat Merged), and across the three national groupings (Canadian, American, and British).

Overall, JG's mean differences between the *caught* and *cot* formant values were 31.7 Hz for F1, and 145.43 Hz for F2. Based on these values, JG would be best placed in the Somewhat Merged group. The differences in JG's formant values for *cot* and *caught* between the early and late parts of the interviews, looking across all interviews, are given in Table 4.

Table 4. JG's overall *cot* and *caught* formant value differences (in Hz), separated into early and late portions of the interviews. Negative values in the change from early to late indicate that formant values became closer over time.

	F1	F2
Early difference caught:cot	31.2	179.1
Late difference caught:cot	33.45	124.12
Difference change from early to late	**2.25**	**-54.98**

[4] In six of the 12 interviews, fewer than 8 *caught* tokens were produced; in one of these interviews, JG produced only one *caught* token in the first half, and in another, he produced only one *caught* token in the second half (only five tokens were produced in total in these two interviews).

Overall, the mean F1 difference between the vowel types remains quite stable throughout the course of the interviews (between 31 and 34Hz). The mean F2 difference, on the other hand, decreases by nearly 55Hz over time, mainly due to a 70Hz increase in the mean F2 of the *caught* tokens over time (see Figure 1in the Supplementary Material available at http://www.linguistics.ubc.ca/people/jcabel).

The formant values for JG's *cot* and *caught* productions with the different merger groups – Merged, Unmerged, and Somewhat Merged – are given in Table 5.

Table 5. *Cot* and *caught* formant values differences (in Hertz) for JG for each merger group. Negative values in the overall change from early to late indicate that formant values became closer over time.

	Merged		**Unmerged**		**Somewhat Merged**	
	F1	**F2**	**F1**	**F2**	**F1**	**F2**
Early difference cot-caught	8.56	152.24	29.38	63.18	53.04	222.88
Late difference cot-caught	12.25	120.23	20.305	121.62	45.5	133.57
Change in cot-caught difference from early to late	**3.69**	**-32.01**	**-9.07**	**58.4**	**-7.54**	**-89.31**

While JG's F1 difference with the Merged group is fairly small, it increases slightly over time; and while his F2 difference decreases over time, it remains quite large. With the Unmerged group, while JG's F2 difference increases over time, his F1 difference decreases. With the Somewhat Merged group, there are decreases in both F1 and F2 over time: a small one in F1, and a large one – nearly 90 Hz – in F2. This large F2 difference change is primarily due to an increase in the F2 of *caught* of 116Hz from early to late, an increase which is similar to (in fact, larger than) the change in the F2 of *caught* seen in the overall results. (This change is shown in Figure 2 of the Supplementary Material available at http://www.linguistics.ubc.ca/people/jcabel).

The formant values for JG's *cot* and *caught* productions with the different national groups – American, British, and Canadian – are given in Table 6.

Table 6. *Cot* and *caught* formant value differences (in Hertz) for JG for each national group. Negative values in the overall change from early to late indicate that formant values became closer over time.

	American		**British**		**Canadian**	
	F1	**F2**	**F1**	**F2**	**F1**	**F2**
Early difference cot-caught	29.36	226.165	34.44	73.67	17.47	152.29
Late difference cot-caught	14.17	115.74	38.725	82.89	45.99	169.23
Change in cot-caught difference from early to late	**-15.19**	**-110.425**	**4.285**	**9.22**	**28.52**	**16.94**

With the British group – three Unmerged speakers and one Somewhat Merged speaker – his F1 and F2 differences increase slightly over time. With the Canadian group – two Merged speakers and two Somewhat Merged speakers – JG's differences again increase over time. With the American group – which includes one Merged speaker, one Unmerged speaker, and two Somewhat Merged speakers – JG's F1 and F2 differences decrease over time: by a relatively small amount in the case of F1, and by a large amount – 110Hz – in the case of F2. This F2 difference change is due primarily to an increase in the F2 of *caught* of nearly 150Hz over time. This change is shown in Figure 3 (see the Supplementary Material available at http://www.linguistics.ubc.ca/people/jcabel).

A post-hoc analysis was conducted to explore the observed increases in F2 from early to late for the *caught* variable with the overall, American, and Somewhat Merged groups. It was found that the type of coda following the *caught* or *cot* vowel had a dramatic impact on the vowel formants, even at the 50% mark; formant values before sonorant consonants were lower than those before obstruent consonants or in word-final coda-less syllables. The greatest effect was found on the F2 of JG's *caught* vowels: overall, the mean F2 of *caught* was 822Hz before sonorant consonants, and 1027Hz in all other environments – a difference of 205Hz. By comparison, the mean F2 of JG's *cot* tokens was 1028 Hz before sonorant consonants, and 1135Hz in all other environments – a difference of 107 Hz. The overall sonorant coda vs. non-sonorant coda changes are shown in Figure 4 (see the Supplementary Material available at http://www.linguistics.ubc.ca/people/jcabel). The proportion of sonorant consonant codas was highest in the early productions of *caught*: 48% (20 of 42) early *caught* tokens were pre-sonorant, compared to 28% of late *caught* tokens (16/57), 36% of early *cot* tokens (42/115), and 33% of late *cot* tokens (32/99). The highest proportion of pre-sonorant *caught* tokens among the three merger groupings was found in the early productions of the Somewhat Merged group (14/23, or 61%), and the highest proportion of pre-sonorant *caught* tokens among the three regional groupings was found in the early productions of the American group (15/23, or 65%). It appears that it is the proportion of sonorant consonants that is driving the change in F2 of the *caught* tokens, rather than a particular interlocutor grouping.

3.2 Canadian Raising

JG's number of tokens of the [aʊ] diphthong in CR and not-CR environments was quite small: an average of 10 CR tokens per interview (maximum in one interview 18, minimum 3) and 7.92 notCR tokens per interview (maximum in one interview 12, minimum 1). The number of tokens in these categories was less than desirable for within-interview analysis, both for JG and for his interlocutors (average number of guest tokens: 17.58 CR, 17.17 notCR). As well, the majority of CR tokens were found in the word <about> (86 of 120 total tokens); a large proportion of these were highly reduced and consequently very schwa-like. Any measurements of changes in the diphthongal qualities of [aʊ] in CR and notCR environments would thus become unreliable. For this reason, [aʊ] was excluded from further analysis.

In the case of [aɪ], the 12 interviewees were grouped according to two characteristics of their productions. The first was the difference between the F1 of the [a] component of the diphthong in CR and not-CR environments (see section 2.2), measured at the point in the vowel of maximum F1. Those talkers who were labeled as Raisers had a 100Hz or greater difference between their CR and non-CR F1 values at either the 25% or 50% marks of the vowel (see section 2.3). In addition, Moreton and Thomas (2007) and Boberg (2008) noted significant fronting of [aɪ] in CR environments, leading to a noticeable increase in F2. Following Moreton and Thomas (2007), the percentage difference between F2 values in CR and not-CR environments was used rather than a Hertz value. Those talkers who were labeled as Raisers had a 9% or greater difference between their CR and non-CR F2 values at either the 90% or 95% mark of the vowel (see section 2.3). Thus, Raisers had large differences in both F1 and F2 between CR and not-CR environments, while Non-Raisers did not.[5] The resulting groupings are given in Table 8 for the F1 differences, and in Table 9 for the F2 differences.

Table 8. Groupings of interviewees by presence of Canadian Raising pattern in F1 for [aɪ] diphthongs. Differences are in Hertz. Negative values indicate that CR formant values are lower than notCR formant values.

Raising group		**Non-Raising group**	
Guest	**Max F1 difference (Hz)**	**Guest**	**Max F1 difference (Hz)**
AM	-163.4	**CL**	-124
BC	-108	**IM**	-28.88
BM	-121.2	**KF**	-23.67
HW	-150.23	**ML**	-13.26
SM	-141.99	**NL**	-1.49
SP	-143.17	**PC**	-4.02

[5] Some previous work has attempted to set benchmarks for a minimum F1 difference in CR (e.g., Boberg 2008, Labov et al. 2006); no reason is given for setting them at a particular value. Moreton and Thomas (2007) do not use a benchmark for either F1 or F2 and do not report a particular pattern for F1 differences, but note that their raisers have a 7% or more difference between their F2 values in CR and notCR environments.

Table 9. Groupings of interviewees by presence of Canadian Raising pattern in F2 for [aɪ] diphthongs. Differences are in percentages. Negative values indicate CR formant values are lower than notCR formant values.

Raising group		**Non-Raising group**	
Guest	**Max F2 difference (%)**	**Guest**	**Max F2 difference (%)**
AM	9.72	**CL**	3.07
BC	9.54	**IM**	14.99
BM	21.69	**KF**	16.19
HW	14.6	**ML**	3.75
SM	20.15	**NL**	8.59
SP	12.99	**PC**	7.28

The total number of tokens of the [aɪ] diphthong in CR and not-CR environments for JG are given in Table 10 (average numbers for guests per interview: 33.5 CR, 53.08 notCR). Again, these numbers were deemed too small for within-interview analysis of changes in JG's productions.

Table 10. Number of tokens of [aɪ] in CR and not-CR environments for JG

Number of tokens	**CR**	**notCR**
Early	70	110
Late	70	105
Total	140	215
Average per interview	11.67	17.92
Maximum in one interview	23	36
Minimum in one interview	6	7

Overall, JG's mean maximum F1 for [aɪ] diphthongs in CR environments was 42.53 Hz lower than that in not-CR environments, and his mean maximum F2 in CR environments was 14.36% higher in than in not-CR environments). Based on these values, JG seems to be best placed in the Non-Raising category. The maximum differences in JG's F1 values in CR and non-CR environments between the early and late parts of the interviews with the various guest groupings (all guests, Raising and Non-Raising, and national groupings) are given in Table 11.

Table 11. JG's CR-notCR F1 differences (in Hz) in early and late parts of the interviews with various guest groups.

Interlocutor Group	**F1 Difference Early**	**F1 Difference Late**	**Change Over Time**	**Direction of Change**
All Guests	41.9	42.58	0.68	Increase
Raising Guests	60.19	44.55	-15.64	Decrease
Non-Raising Guests	23.25	41.61	18.36	Increase
Canadian Guests	57.7	44.33	-13.37	Decrease
Non-Canadian Guests	35.07	43.09	8.02	Increase
American Guests	44.15	61.81	17.66	Increase
British Guests	27.26	22.83	-4.43	Decrease

Overall, JG's F1 difference shows a very slight increase over time. With the Raising and Non-Raising groups, a pattern of divergence is seen: with the Raising group, JG's maximum difference between CR and not-CR decreases over time, while with the Non-Raising group, the maximum difference between CR and not-CR increases over time. The same divergent pattern is seen with the Canadian and Non-Canadian groups: with the Canadian group (4 guests), JG's CR-notCR difference decreases over time, as it did with the Raising group; with the Non-Canadians (8 guests), as with the Non-Raisers, the difference increases over time. Splitting the Non-Canadian group into its constituent subgroups, the difference between JG's CR and not-CR productions with Americans (three of whom are Non-Raisers) increases over time; in contrast, with the British group (three of whom are Non-Raisers), the difference between CR and not-CR environments decreases slightly from the early to late portions of the interviews.

The maximum percentage differences in JG's F2 values in CR and non-CR environments between the early and late parts of the interviews with the various guest groupings (all guests, Raising and Non-Raising, and national groupings) are given in Table 14.

Table 14. JG's CR-notCR F2 differences (in %) in early and late parts of the interviews with various guest groups.

Interlocutor Group	**F2 % Difference Early**	**F2 % Difference Late**	**Direction of Change Over Time**
All Guests	11.64	15.98	Increase
Raising Guests	17.54	14.67	Decrease
Non-Raising Guests	6.12	17.01	Increase
Canadian Guests	20.31	17.2	Decrease
Non-Canadian Guests	7.56	15.49	Increase
American Guests	8.25	14.44	Increase
British Guests	6.72	16.37	Increase

Overall, again, JG's F2 difference increases over time. With the Raising group, as with F1, the difference decreases over time, although it is still quite large in both the early and late portions of the interviews. In contrast, in his productions with the Non-Raising group, there is a much greater increase in the CR versus not-CR F2 difference over time than there is in the F1 differences; the percentage difference more than doubles between early and late portions of the interviews. With the Canadian group, JG's absolute and percentage differences decrease over time just as they did with the Raising group, although the percentage differences are again still quite large in both the early and late portions of the interviews. In his productions with the Non-Canadian group, as with the Non-Raising group, much greater increases occur in JG's F2 differences between CR and not-CR environments than occur in F1. JG's patterns with both the American and British groups are similar to those with the overarching Non-Canadian group and with the Non-Raising group; the percent difference between F2 in CR and notCR environments in the early portions of the interviews is between 6 and 8%, but increases to a double-digit difference by the late portions of the interviews with both groups.

4 Discussion

This study examined two phonetic variables – the *cot-caught* merger and the diphthongs involved in Canadian Raising – as produced by one Canadian English talker, the radio host Jian Ghomeshi (JG), through a series of radio interviews for evidence of spontaneous phonetic imitation with those variables. While spontaneous imitation of a range of phonetic variables is frequently seen in the laboratory or other controlled settings, the content, vocabulary, and speech styles used in the interviews analyzed here were not controlled, raising the question of whether imitation would be found in this type of environment. In the end, across 12 interviews, no evidence was found suggesting that systematic imitative behaviour was driving the changes seen in JG's productions of either the *cot-caught* vowels or the Canadian Raising diphthong [aɪ]. This does not mean that spontaneous phonetic imitation does not occur in non-laboratory interactions. While it may ultimately be the case that true spontaneous phonetic imitation only occurs with low-frequency words produced with a certain shortness of latency between exposure and production, it may also be the case that it does occur outside the laboratory but simply was not found with these variables in this corpus with this talker. In the second instance, the lack of imitation could be due to a number of factors, including the type of interaction which was chosen (i.e., using in-person interviews instead of panel discussions, studio-to-studio interviews, or conversations with regular guests on the program), the general noisiness of the data in the corpus, personal or professional (i.e., in his role as radio host) characteristics of JG's speech or interactional behaviours, the division of JG's attention between multiple facets of the situation (limiting how much attention he can pay to the guest's speech; cf. Abel et al. 2011), the particular variables which were chosen, or the limited amount of data available for these variables in the corpus.

In the case of the *cot-caught* vowels, there is no evidence that the changes seen are the result of either imitation or accommodation; no regular changes were seen in JG's F1 with any of the merger or dialect groups, and the changes in the F2 of the *caught* vowel over time cannot be linked to any particular group. Instead, it appears to be the phonetic environment – in particular, whether or not the segment following the *caught* vowel is a sonorant consonant – which is creating the observed variation in F2, and quirks of distribution – more pre-sonorant *caught* tokens with certain speakers than with others, more pre-sonorant *caught* tokens in early portions of the interviews than late – which result in the observed changes over time. This does not mean that evidence of imitation could not be found if more data were collected to even out the distributional clumping that resulted in the patterns seen here. However, while merged vowels are clearly important in distinguishing dialects (see Labov et al. 2006, Wells 1982 *inter alia*), it may be the case that they are not the best segments to examine for imitative behaviour. For example, Evans and Iverson (2007) found that young adults from a northern English town, whose native dialect has an [ʊ]-[ʌ] merger, did not unmerge those vowels even after two years of regular exposure to an unmerged dialect (Southern Standard British English). If changes to a merger do not occur after a period of several years, it may be optimistic to expect them to occur over the course of an 18-minute interview.

In the case of [aɪ], again, there is no evidence that JG is systematically imitating the Canadian Raising patterns of his interlocutors; in fact, divergent behaviour is found more often than convergent behaviour. With the Raising group, the difference between JG's formant values in raising and non-raising environments decreases for both F1 and F2; the same pattern holds with the Canadian group, who are all raisers. With the Non-Raising group, on the other hand, the difference between formant values increases for both F1 and F2; the same pattern holds with the Non-Canadian group (75% of whom are non-raisers). Within the Non-Canadian group, JG displays a consistent pattern of increasing formant value differences over time with the American group, in which only one of four guests is a raiser. This pattern is not found with the other Non-Canadian subgroup, the British group, which also has one raising member.

While imitation does not appear to be behind this pattern with the Americans, accommodation could be. JG's divergence from the predominant American pattern of non-raising may be a way of establishing group membership (Giles and Ogay 2007): in particular, of establishing JG's 'Canadian-ness' – or 'CBC-ness' – versus 'American-ness'. The distinctions between Canadian and American English, and the distinction of Canadian English as an entity separate from American English, are not unknown or unimportant to Canadians: indeed, the phonetic features of Canadian English as distinct from American English are quite regularly discussed in the media (for a recent example, see Howell 2011) and exaggerated for entertainment value (e.g., the *Royal Canadian Air Farce* character 'Mike [mʌɪk] from Canmore', or the recent Canadian beer commercial in which an American who taunts a Canadian with the phrase *no doot aboot it* [no doubt about it] is pummelled for his trouble). Indeed, given that part of the mandate of the CBC is to "serve the needs and interests and reflect the circumstances and aspirations, of Canadian men, women and children" (Canadian Broadcasting Corporation 1994: Policy 1.1.1, d(iii)), if this distinction is important to Canadians, it may also be important to the CBC as a whole. As for why JG displays more Canadian Raising with American guests than with British guests, one possibility is that "[s]peakers of Canadian English generally define their dialect with respect to American English, and not with respect to British English" (Gold and Tremblay 2006: 257). Another, more individual-oriented possibility is that as JG spent the early part of his life in Great Britain, he may personally feel more affinity with those guests and not feel the need to distinguish himself from them linguistically. In terms of JG's raising behaviour with Canadian guests – i.e., the decrease in the difference between the formants in raising and non-raising environments over time – the lack of convergence could be due to a number of factors, such as a lack of a need for convergence (i.e., because JG is already established as part of the Canadian/CBC ingroup, he does not need to re-establish himself), or a desire not to over-accommodate (see Giles and Ogay 2007) to Canadian speakers (e.g., a noticeable increase in CR on JG's part may make Canadian guests uncomfortable about their linguistic behaviour). In order to determine if accommodation is a plausible driver of JG's raising behaviour, two things will be necessary: a larger sample of both [aɪ] tokens and interlocutors, to determine if the observed changes are reliable, and information about his self-identification with various groups (e.g., Canadians, British-born Canadians, CBC employees), to determine if the observed changes align with his identification with those groups.

What would we expect if the variation seen is not being driven by either imitation or by accommodation, but by audience design concerns? Bell (1984) suggests that the consumers of radio programming in particular are more like bystanders to a conversation – true audience members – rather than absent, judgmental referees: "announcers rely solely on their speech to project whatever relationship they have with the audience. Their style draws its effect from the norms of who such a style is addressed to in face-to-face interaction. They use style as an expressive instrument, a declaration of identity, saying to the audience 'you and I are ingroup'" (192). If this is the case, then either or both of the results seen with the two variables – the apparent lack of socially-driven change with the *cot-caught* vowels, and the divergent behaviour with the [aɪ] diphthong – could be due to audience design. If JG holds his relationship to his audience paramount even when he is talking to an in-studio guest, then we would not necessarily expect him to make any changes based on who his immediate interlocutor is; because he and his audience are ingroup, if he maintains his usual speech pattern – as he does with his *cot-caught* vowels – it is to maintain solidarity with that ingroup, rather than changing his productions to attempt to create solidarity with the interviewee. On the other hand, if the Canadian English ingroup has a widespread desire to distinguish itself linguistically from American English speakers in particular, then JG, as part of this ingroup, may diverge phonetically from American interlocutors in particular, as seen in his variation with the [aɪ] diphthong. To fully explore these possibilities, it would again be important to collect more data, as well as to gather data about the usual listening audience for *Q* –and about JG's knowledge of the composition of his audience – to determine to what degree JG's productions reflect those of the audience.

This study employed an untried means of data collection: using publicly-available podcasts of (relatively) naturally occurring conversation. This method allowed for the cost-free compilation of a small corpus, with the possibility of substantial expansion given the archive of programs available through the CBC website. Similar archives are available from a number of public broadcasters, including the Australian Broadcasting Corporation (http://www.abc.net.au/services/podcasting/), the British Broadcasting Corporation (http://www.bbc.co.uk/podcasts), Radio New Zealand (http://www.radionz.co.nz/podcasts), and NPR (http://www.npr.org/rss/podcast/podcast_directory.php). Podcasts are also available from a wide variety of other sources, both professional and amateur. In addition to providing a wealth of data sources, using podcasts also allows for the circumvention, at least to some degree, of the observer's paradox. While it is true that the participants in these interviews knew that they were being listened to – in fact, they would likely have been disappointed had they not been listened to – there was no expectation that what they said would be mined for phonetic data.

However, in taking the study of spontaneous phonetic imitation out of the confines of the laboratory and into the realm of 'real-world' conversational interaction – through podcasts or otherwise – it is apparent that what may be fairly straightforward in a controlled setting is often less so in an uncontrolled one. For example, if one hopes to study changes in talkers' [aʊ]-diphthong productions over time in the laboratory, it is relatively simple to create a wordlist (controlled for word frequency and familiarity) with the desired number of tokens in the desired range of environments and have the talkers repeat those tokens a desired number of times at the desired temporal latency (i.e., right after hearing the word or at some later time), or to create a dyadic interaction task which requires talkers to use words with [aʊ] diphthongs to accomplish their task. Granted, not all talkers in an experiment will produce the same amount of usable data (due to errors in completing the task, fatigue, boredom, etc.), but it is often possible to make a reasonable guess at how many talkers will be required to collect the desired amount of data. In a study of the type described here, however, it is nearly impossible to predict in advance how many tokens of the predetermined desired types will be available from any given talker, whether they will be distributed evenly over the time period under examination, or whether the productions will be sufficiently like their citation forms to display the canonical features of those variables. Ultimately, it may be the case that future studies in this vein are better served by taking a different approach to determining what will be examined. For example, the corpus could be better tailored to looking at a particular variable – e.g., imitation of the [aɪ] diphthong – by selecting interviews in which words with [aɪ] diphthongs like <write>

are likely to occur frequently, such as interviews with writers or songwriters. This is similar to the approach taken by Hall-Lew et al. (2010), who examined the pronunciation of <Iraq> in political speeches (although their study did not look at imitation *per se* or at changes over time). Alternately, determining the overall range of data available in a corpus and focusing on those variables which are best represented, rather than attempting to pre-set which variables will be examined, may help to avoid some of the issues mentioned above.

As well as using corpus data, it may also be useful to further imitation research to make the laboratory environment more like that found in natural conversational settings. Ongoing research by Abel et al. (2011) has begun to look at the effect on spontaneous phonetic imitation of dividing participants' attention between tasks (e.g., having them answer math questions while listening to a model talker). Other aspects which could be varied include the length of phrases participants hear, whether they hear all words/phrases the same number of times, and whether the same amount of time elapses between hearing the word/phrase and producing it. Whether corpora of conversations or less artificial laboratory studies are used, the insights gained from using richer and more complex speech situations will be important in furthering our understanding of imitative phenomena in speech.

References

Abel, Jennifer, Molly Babel, and Alexis Black. 2011. Phonetic imitation in contexts of stimulus-directed and non-stimulus-directed attention. Poster presented at the 162nd Meeting of the Acoustical Society of America, San Diego, CA, November 2011.

Abrego-Collier, Carissa, Julian Grove, Morgan Sonderegger, and Alan C. L. Yu. 2011. Effects of speaker evaluation on phonetic convergence. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 192-195.

Babel, Molly. 2010. Dialect divergence and convergence in New Zealand English. *Language in Society* 39: 437-456.

Babel, Molly. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40: 177-189.

Bell, Allan. 1984. Language Style as Audience Design. *Language in Society* 13(2): 145-204.

Bilous, Frances R., and Robert M. Krauss. 1988. Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language and Communication* 8(3/4): 183-194.

Boberg, Charles. 2008. Regional phonetic differentiation in standard Canadian English. *Journal of English Linguistics* 36(2): 129-154.

Boersma, Paul, and Weenink, David. 2009. Praat: doing phonetics by computer (version 5.1.17) [computer program]. Retrieved 26 September 2009 from http://www.fon.hum.uva.nl/praat/. Institute of Phonetic Sciences, Amsterdam.

Brown, Mitchell. 2003. Diversity key for CBC host. *Thornhill Leader*, October 1, 2003. Retrieved 30 July 2011 from http://jian.ca/press/diversity-key-for-cbc-host/.

Canadian Broadcasting Corporation. 1994. CBC/Radio-Canada Mandate (CBC/Radio-Canada Corporate Policy 1.1.1). Retrieved 20 March 2012 from http://www.cbc.radio-canada.ca/docs/policies/program/mandate.shtml.

Canadian Broadcasting Corporation. 2011. *Q: About the Show*. Retrieved 30 July 2011 from http://www.cbc.ca/q/about/.

Chambers, J.K. 1989. Canadian Raising: Blocking, Fronting, etc. *American Speech* 64(1): 75-88.

Chambers, J.K. 2006. Canadian Raising retrospect and prospect. *Canadian Journal of Linguistics* 51(2/3): 105-118.

Coupland, Nikolas. 1980. Style-shifting in a Cardiff work-setting. *Language in Society* 9(1): 1-12.

Coupland, Nikolas. 2001. Dialect stylization in radio talk. *Language in Society* 30: 345-375.

Evans, Bronwen G., and Paul Iverson. 2007. Plasticity in vowel perception and production: a study of accent change in young adults. *Journal of the Acoustical Society of America* 121(6): 3814-3826.

Giles, Howard. 1973. Accent mobility: a model and some data. *Anthropological Linguistics* 15(2): 87-105.

Giles, Howard, Donald M. Taylor, and Richard Bourhis. 1973. Towards a theory of interpersonal accommodation through language: some Canadian data. *Language in Society* 2: 177-192.
Giles, Howard, and Tania Ogay. 2007. Communication Accommodation Theory. In *Explaining Communication: Contemporary Theories and Exemplars*, Bryan B. Whaley and Wendy Samter (eds.), 293-310. Mahwah, NJ: Lawrence Erlbaum Associates.
Gold, Elaine, and Mireille Tremblay. 2006. *Eh?* and *Hein*?: Discourse particles or national icons? *Canadian Journal of Linguistics* 51(2/3): 247-263.
Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2): 251-279.
Gregory, Stanford W. 1990. Analysis of fundamental frequency reveals covariation in interview partners' speech. *Journal of Nonverbal Behavior* 14(4): 237-251.
Gregory, Stanford W., Stephen Webster, and Gang Huang. 1993. Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language and Communication* 13(3): 195-217.
Gregory, Stanford W., and Stephen Webster. 1996. A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology* 70(6): 1231-1240.
Gregory, Stanford W., and Timothy Gallagher. 2002. Spectral analysis of candidates' nonverbal vocal communication: predicting U.S. presidential election outcomes. *Social Psychology Quarterly* 65(3): 298-308.
Hall-Lew, Lauren, Elizabeth Coppock, and Rebecca L. Starr. 2010. Indexing political persuasion: Variation in the *Iraq* vowels. *American Speech* 85(1): 91-102.
Hay, Jennifer, Stefanie Jannedy, and Norma Mendoza-Denton. 1999. Oprah and /ay/: Lexical frequency, referee design and style. In *Proceedings of the 14th International Congress of Phonetic Sciences*, John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville and Ashlee C. Bailey (eds.), 1389-1392.
Howell, Peter. 2011. Second thoughts "a-boat" how Canadians pronounce "about". *Toronto.com*, June 20, 2011. Retrieved 21 March 2012 from http://www.toronto.com/blog/post/689807--second-thoughts-a-boat-how-canadians-pronounce-about.
Joos, Martin. 1942. A phonological dilemma in Canadian English. *Language* 18(2): 141-144.
Kim, Midam, William S. Horton, and Ann Bradlow. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology* 2: 125-156.
Labov, William, Sharon Ash, and Charles Boberg. 2006. *Atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.
Moreton, Elliott, and Erik R. Thomas. 2007. Origins of Canadian Raising in voiceless-coda effects: A case study in phonologization. In *Laboratory Phonology 9*, Jennifer Cole and José I. Hualde (eds.), 37-64. Berlin: Walter de Gruyter.
Namy, Laura L., Lynne C. Nygaard, and Denise Sauerteig. 2002. Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology* 21(4): 422-432.
Natale, Michael. 1975. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology* 32(5): 790-804.
Nielsen, Kuniko. 2011. Specificity and abstraction of VOT imitation. *Journal of Phonetics* 39: 132-142.
Pardo, Jennifer S. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119(4): 2382-2393.
Shockley, Kevin, Laura Sabadini, and Carol A. Fowler. 2004. Imitation in shadowing words. *Perception and Psychophysics* 66(3): 422-429.
Sjölander, Kåre, and Jonas Beskow. 1996. Wavesurfer (version 1.8.5) [computer program]. Retrieved 5 January 2010 from http://www.speech.kth.se/wavesurfer/. Department of Speech, Music and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden.
Wells, J.C. 1982. *Accents of English* (3 vols.). Cambridge: Cambridge University Press.

The temporal interpretation of modals in Mandarin Chinese[*]

Sihwei Chen
University of British Columbia

This paper investigates how modals anchor the temporal perspective and the temporal orientation in Mandarin. Mandarin exhibits distinct temporal orientations for modals with different types of conversational background (Kratzer 1991), epistemic vs. circumstantial (Condoravdi 2002, Stowell 2004, Werner 2006, van de Vate 2010, Matthewson to appear, among others). Moreover, both types of modals are found to allow shifting of temporal perspective from the speech time. Based on the (in)compatibility of the modals and aspectual/temporal marking, I propose that epistemic modals do not encode temporal orientation, instead relying on the aspectual and temporal marking on their prejacents. Circumstantial modals, on the other hand, consistently encode futurity in their semantics (Enç 1996, van de Vate 2010, Abusch to appear, among others).

1 Introduction

The interaction of modals and time has been elaborated in many recent crosslinguistic studies (e.g. Enç 1996, Condoravdi 2002, Stowell 2004, Werner 2006, Borgonovo and Cummins 2007, Demirdache et al. 2008, Laca 2008, Matthewson to appear, among others). In the sense of Condoravdi (2002), the temporal interpretation of modals involves two notions, *temporal perspective* and *temporal orientation*. *Temporal perspective* is the time at which the modal background, the evidence available or the laws in effect, is accessed, whereas *temporal orientation* is the relation between the temporal perspective and the time of the prejacent (i.e., the proposition embedded under the modal). For example, the English epistemic modal *might* in (1) concerns the possibility in view of speaker's knowledge at speech time as to a state of various affairs located in the future, present, and or past: it has a present T.P. and allows an eventive prejacent to have a future T.O. (1a) and a stative prejacent a present or future T.O. (1b); with help of the perfect *have*, both prejacents show a past T.O. (1c-d).

(1) a. He might get sick tomorrow/??now/*yesterday.
b. He might be sick tomorrow/now/*yesterday.
c. He might have gotten sick *tomorrow/*now/yesterday.
d. He might have been sick *tomorrow/*now/yesterday. (Condoravdi 2002: 60)

This paper investigates how modals anchor the temporal perspective and the temporal orientation in Mandarin. Mandarin has a very rich aspectual system, which Smith and Erbaugh (2005), as well as Lin (2006), has argued to be the resource of temporal reference. Various other literature has also linked viewpoint to certain types of modality. Ren (2008) argues that circumstantial modals (her deontic and dynamic modals) in Mandarin encode futurity, whereas epistemic modals diverge as to whether they expand forward the topic time of the eventuality to the future. However, the modals she classifies as epistemic modals actually include both circumstantial modals and future modals. Also, the eventuality types, which are a factor influencing temporal interpretation, are not kept uniform under the modals in her analysis. Considering these drawbacks, I modify Ren's (2008) study by controlling both modality types and eventuality types with respect to overt temporal or aspectual marking within the prejacent clause.

The methods of eliciting sentences follow Matthewson (2004). The judgments for the temporal interpretations are obtained by offering the sentences containing modals to consultants in different discourse contexts. For instance, a sentence which is accepted in a modal context with a future T.O. is

[*] I gratefully thank Lisa Matthewson, Hotze Rullmann, my QP committee, and the LING 530C class at UBC. Thanks also to the audience at the QP mini-conference on April 17th, 2012 and the modality workshop at Ottawa on April 20-21th, 2012. This research was supported by SSHRC grant #410-2011-0431. Any remaining errors are mine.

judged to be true in that context. If a sentence is rejected in a particular situation, follow-up elicitation including direct translation and grammatical judgments is supplemented to inquire if the sentence is false or infelicitous in that situation. The meta-language used to elicit the sentences in question is Mandarin.

This paper is organized as follows. Section 2 presents classification of Mandarin modals based on Kratzer (1991) and Rullmann et al. (2008). Section 3 introduces Mandarin temporal system by reviewing Lin's (2006) analysis. Section 4 discusses temporal orientation possibilities of the modals and how they shift temporal perspective. A split analysis of the different temporal orientation between circumstantial and epistemic modals is given in Section 5. Section 6 discusses consequences of the analysis and concludes this paper.

2 Classification of modals in Mandarin

Kratzer (1991) argues that uses of modals are decided by an implicit conversational background in context, which consists of two sets of propositions: the modal base and the ordering source. The modal base maps each world onto a set of accessible worlds, over which the modal quantifies, and the ordering source ranks and restricts the domain of the accessible worlds. In terms of Kratzer, epistemic modals have an epistemic modal base and can come with a stereotypical or doxastic ordering source. Circumstantial modal base can also combine with different ordering sources depending on the context, such as a deontic, bouletic, teleological or empty ordering source. Rullmann et al. (2008) adopt Kratzer's (1991) conversational backgrounds but propose that languages might vary in whether they have a contextually given or lexicalized conversational background.[1] For instance, the St'át'imcets (Lilloet Salish) modals lexically specify quantificational strength but allow variable conversational backgrounds. The Pesisiran Javanese modals have specified quantificational strength as well as a selective modal base (Vander Klok 2008). I follow the emerging theoretical typology and give a Kratzer-style analysis of Mandarin modals. There's an identifiable class of auxiliary verbs that express modality in nature, despite the imperfect criteria for distinguishing auxiliaries, verbs, and adverbs (cf. Chao 1968, Li and Thompson 1981, Tsang 1981, Lin and Tang 1995, Li 2004, Ren 2008 etc.). I classify five frequently discussed modals in this paper: *yīdìng, kěnéng, bìxū*, *kěyǐ*, and *yīnggāi*.[2] The five modals can be analyzed as lexically encoding both two main types of modal bases (epistemic vs. circumstantial) and the quantificational strength (necessity or possibility):

(2) a. Context: *He looks for his dog all over the house, but cannot find it.*
xiǎogǒu *yīdìng*/**kěnéng* pǎodiào le.[3]
small.dog EPIS.NEC/EPIS.POS run.away PRF
'The dog must have escaped.'

b. Context: *He was playing with a dog. When he just turned, he couldn't find the dog.*
xiǎogǒu **yīdìng*/*kěnéng* pǎodiào le.
small.dog EPIS.NEC/EPIS.POS run.away PRF
'The dog may have escaped.'

(3) a. Context: *My meeting with the student isn't done but he has to attend a class.*
wǒ *bìxū*/**kěyǐ* rang tā líkāi.
I CIRC.NEC/CIRC.POS let he leave
'I have to let him go.'

b. Context: *Only family members are allowed to enter the patient's room during visiting hours, but you're exceptional since you are a really close friend.*

[1] They also differ from Kratzer in adopting choice function but not the ordering source to restrict the domain of the accessible worlds universally quantified over.
[2] There are two more modals which often overlap with *bìxū* and *kěyǐ* respectively: *dé* and *néng(gòu)*. Like *bìxū* and *kěyǐ*, they lexically encode a circumstantial modal base but may differ from *bìxū* and *kěyǐ* in allowing different ordering sources. The reader who is interested in their use can refer to Tsang (1981) and Li (2004).
[3] Abbreviations used in morpheme glosses are as follows: ABIL = ability, CIRC = circumstantial, CL = classifier, EPIS = epistemic, EXP = experiential, FUT = future, LOC = locative, NEC = necessity, NEG = negative, PFV = perfective, PL = plural, POS = possibility, PRF = perfect, PROG = progressive, PRT = particle, REL = relative clause, W = weak.

nǐ *bìxū/kěyǐ jìnlái.
you CIRC.NEC/CIRC.POS come.in
'You may come in.'

The only exception to this generalization is the weak necessity modal *yīnggāi*, which is ambiguous between both types of modal bases (4). The Mandarin modal system is summarized in Table 1.

(4) a. Context: *It looks like my friend has left the party; her bag is gone, but she might have just taken it into the bathroom.*
wǒ péngyǒu *yīnggāi* yǐjīng huíqù le. EPISTEMIC
my friend W.NEC already return PRF
'My friend must have left.

b. Context: *We rotate to do chores and today's your turn.*
nǐ *yīnggāi* xǐ pánzǐ. DEONTIC
you W.NEC wash dishes
'You ought to do the dishes.'

Table 1. The Mandarin modality system

<table>
<tr><td rowspan="2">CB / QF</td><td>epistemic</td><td colspan="4">circumstantial</td></tr>
<tr><td>stereotypical</td><td>bouletic</td><td>deontic</td><td>teleological</td><td>stereotypical</td></tr>
<tr><td>necessity</td><td>yīdìng</td><td colspan="3">bìxū</td><td rowspan="2">huì</td></tr>
<tr><td>weak necessity</td><td colspan="4">yīnggāi</td></tr>
<tr><td>possibility</td><td>kěnéng</td><td colspan="4">kěyǐ</td></tr>
</table>

3 Viewpoint aspect and temporal reference in Mandarin

Before assessing the temporal reference in modality, how modal-less sentences express their temporal reference has to be considered. This section introduces the Mandarin temporal system by reviewing Lin's (2006) analysis. Mandarin Chinese lacks tense morphology but utilizes many other factors such as lexical and viewpoint aspectual morphemes, temporal adverbs and pragmatic reasoning and so forth to determine the temporal reference of sentences (Smith and Erbaugh 2005, Lin 2003, 2006).[4] Lin (2006) argues that viewpoint aspect in Mandarin give both aspect and tense: Sentences with an imperfective aspectual morpheme, like the progressive maker *zài* in (5a), have a present interpretation, whereas sentences with a perfective aspectual morpheme, such as the marker *-le* in (5b), have a past interpretation.

(5) a. tāmen *zài* chàng gē. IMPERFECTIVE: PRESENT
they PROG sing song
'They are singing songs.'

b. tāmen chàng-*le* yī-shǒu gē. PERFECTIVE: PAST
they sing-PFV one-CL song
'They sang a song.'

However, not every sentence can carry an overt viewpoint aspect; for instance, stative verbs and achievement verbs are interpreted as present and past respectively, without any aspectual marker:

(6) a. tāmen hěn cōngmíng. STATIVE: PRESENT
they very smart
'They are very smart.'

b. tā dǎ-pò yī-gè bēizǐ. ACHIEVEMENT: PAST

[4] It has been a debate whether null tense exists in Mandarin but such discussion is beyond the topic of this paper; the interested reader can refer to Hu, Pan & Xu (2001), Sybesma (2007), and Lin (2011).

he hit-break one-CL cup
'He broke a cup.'

Lin incorporates into his analysis Bohnemeyer and Swift's (2004) theory that default aspectual interpretations depend on the telicity of the relevant eventuality, so a stative verb like (6a), which is atelic, has a covert imperfective aspect by default, whereas a telic event realized by a resultative verb compound in (6b), has a covert perfective aspect. The semantics of imperfective and perfective aspect Lin proposes is listed in (7). While both aspects specify the inclusive relationship between the topic time and the event time, the perfective aspect additionally encodes the evaluation time t_0, which is preceded by the topic time.

(7) a. Imperfective aspect =: $\lambda P_{<i,t>}\ \lambda t_{Top}\ \exists t\ [t_{Top} \subseteq t \wedge P(t)]$ (Lin 2006: 4)

b. Perfective aspect =: $\lambda P_{<i,t>}\ \lambda t_{Top}\ \boldsymbol{\lambda t_0}\ \exists t\ [t \subseteq t_{Top} \wedge P(t) \wedge \boldsymbol{t_{Top} < t_0}]$ (Lin 2006: 6)

With a default rule which assigns the speech time as the value of the evaluation time or the topic time at the root level, (7) correctly accounts for sentences with an overt or covert aspect. For instance, (7a) applies to (6a) and the default speech time is included within the situation: $\exists t\ [\boldsymbol{t_{Speech} \subseteq t} \wedge$ smart (they)(t)], thus giving a present interpretation. (7a) also correctly explains a shifting effect brought about by a past temporal adverb: In (8), the past time adverb gives the value of the topic time, $\exists t$ [**before ⊆ t** ∧ smart (they)(t)], giving a past interpretation.

(8) tāmen *yǐqián* hěn cōngmíng. STATIVE: PAST
they before very smart
'They were very smart before.'

In contrast, a past time adverb always agrees with the perfective aspect, (6) vs. (9), as the perfective aspect already encodes a past meaning as part of its semantics via the precedence relation $t_{Top} < t_0$.

(9) tā *zuótiān* dǎ-pò yī-gè bēizǐ. ACHIEVEMENT: PAST
he yesterday hit-break one-CL cup
'He broke a cup yesterday.'

Future is lexically marked by an auxiliary/modal *huì* 'will/would' regardless of whether the following verb is eventive or stative (10). Without *huì*, (10a) with the activity verb only has a habitual reading, and (10b) with the stative verb can only be present tense (cf. (6a)).

(10) a. tāmen *(*huì*) chàng gē.[5]
they FUT sing song
'They will sing songs.'

b. tāmen *(*huì*) hěn máng.
they FUT very busy
'They will be very busy.'

In Lin's analysis, the function of *huì* is to locate the topic time after the evaluation time, as in (11), where the evaluation time t_0 can be the speech time (10) or the matrix event time if *huì* is embedded in a subordinate clause (12).

(11) $[[hui]] = \lambda P_{<i,t>}\ \lambda t_{Top}\ \boldsymbol{\lambda t_0}\ [P(t) \wedge \boldsymbol{t_0 < t_{Top}}]$ (Lin 2006: 18)

(12) tā shuō tā *huì* hěn máng.
he say he FUT very busy
'He said that he would be very busy.'

[5] (10a) is grammatical if *huì* is interpreted as an ability modal ('they can sing songs'), cf. fn. 8.

Lin argues that (11) can explain why *huì* is incompatible with a perfective aspect under its scope (13a) but is compatible with an imperfective marker (13b): The function denoted by *huì* requires an argument of type <i, t>. This condition is only met by the output of the imperfective aspect, but not by the output of the perfective marker, which is of type < i, <i, t>>.

(13) a. tāmen *huì* chàng(*-*le*) gē.
they FUT sing- PFV song
'They will sing songs.'

b. tāmen *huì* *zài* chàng gē.
they FUT PROG sing song
'They will be singing songs.'

The Mandarin temporal system can be summarized in Table 2. Future reference is marked by *huì*, which is compatible with the imperfective aspect. If *huì* is absent in sentences, an (overt or covert) imperfective aspect gives a present interpretation by default and can be shifted to the past by a past time adverb if the verb is stative, whereas a perfective aspect always yields a past time.

Table 2. The Mandarin temporal system

	Past	Present	Future
State	*past time adverb*	*(covert) imperfective*	*huì*
Event	*perfective*	*imperfective*	

4 Temporal interpretation of modals in Mandarin

Having reviewed the temporal interpretations of modal-less sentences, I now examine both the temporal orientation (T.O.) and temporal perspective (T.P.) of epistemic and circumstantial modals in combination with prejacents of different eventuality types. I will show that both epistemic and circumstantial modals in Mandarin allow shifting of T.P. from the speech time but they differ in whether they allow a past T.O.

4.1 Epistemic modals

A stative event under an epistemic modal can happen at the same time as the time at which the modal base is calculated, giving a present T.O. (14), or have happened before the evaluation time of the modal, giving a past T.O. (15). The past T.O. is rendered possible by a past time adverb.

(14) Context: *You can't find your friend anywhere and the last place you have not checked is Mahony's. You're now at Mahony's and see your firend's umbrella outside the door.*
tā yīdìng zài zhèlǐ.
he EPIS.NEC LOC here
'He must be here.'

(15) Context: *You can't find your friend anywhere and the last place you have not checked is Mahony's. You're now at Mahony's and find your firend's umbrella but don't see him around. The bartender says that there was a guy looking like what you describe.*
tā *gānggāng* yīdìng zài zhèlǐ.
he just.before EPIS.NEC LOC here
'He must have been here just before.'

To express a future T.O., the stative prejacent has to take the future marker *huì*:

(16) Context: *You would like to visit your friend tomorrow but you are not sure whether he will be at home. You get information from his sister that their family will have a get-together tomorrow which requires every member to be there.*
tā yīdìng *(*huì*) zài jiā.

he   EPIS.NEC   FUT   LOC   home
'He will be at home.'

On the other hand, an eventive verb embedded under an epistemic modal requires the imperfective aspect for a present T.O. (17) and the perfective aspect for a past T.O. (18). For a future T.O., again *huì* is obligatory (19). These sentences are ungrammatical without the aspectual or future marking.

(17) Context: *You hear the uproar and clink of bottles from the living room.*
tāmen   kěnéng   *(*zài*)   hē   jiǔ.
they   EPIS.POS   PROG   drink   wine
'They may be drinking.'

(18) Context: *Your brothers come very late, uttering ravings and stinking of wine.*
tāmen   yīdìng   hē*(-*le*)   jiǔ.
they   EPIS.NEC   drink-PFV   wine
'They must have drunk wine.'

(19) Context: *You don't see your brother at dinner time. Your mother tells you that he just had a quarrel with his girlfriend before you came back and said he was going for a walk alone. You know he always likes to drink when he's in bad mood.*
tā   yīdìng   *(*huì*)   qù   hē   jiǔ.
he   EPIS.NEC   FUT   go   drink   wine
'He will go drinking.'[6]

The temporal orientations of the epistemic modals exhibit exactly the same patterns as the temporal interpretations of unembedded sentences, in which a non-future T.O. heavily relies on the interaction of viewpoint aspect and eventuality type, whereas a future T.O. is acquired only when *huì* is present. The fact is not surprising as the T.O. of a modal is essentially the relation between the evaluation time at which modal background is accessed and the event time of the prejacent. T.O. thus encodes the same temporal relationship as a perfective viewpoint aspect and the future auxiliary *huì* (cf. Section 3). In other words, a past T.O. for epistemic modals correlates with the semantics of a perfective aspect and a future T.O. correlates with the semantics of the future auxiliary.

In the contexts presented so far, the temporal perspective is always located at the speech time. The speaker makes an inference as to the truth of the prejacent based on his knowledge at the time at which he speaks. Logically, the speaker can also evaluate the prejacent based on his knowledge prior to the speech time, giving a past temporal perspective, but a past T.P. for epistemic modals is not commonly acknowledged in the literature (e.g. Condoravdi 2002, Stowell 2004, Hacquard 2009, Demirdache and Uribe-Etxebarria 2008, Laca 2008 etc.) due to the fact that the epistemic state of speakers is in general accessed at the speech time. However, with a rich context, epistemic modals in various languages are found to allow a past T.P. without involving indirect speech or an elided attitude verb (Eide 2003, Matthewson to appear, Matthewson and Rullmann 2012). Interestingly, Mandarin demonstrates that shifting of the T.P. to the past is possible for epistemic modals without extra lexical or morphosyntactic resources. Below are examples of the past T.P. with a present (20), past (21) and future T.O. (22):

(20) Context: *Pat is going on a long trip and Stacey promises to feed his pet for him while he's gone. When Stacey headed to the pet store, she found she actually didn't know what kind of animal Pat has. She remembered Pat likes dogs so she bought a bone. However, 'it's a snake!' she realizes when she comes to Pat's house. Pat asks her, 'Why did you buy a bone?' Stacey says,'Well, you might have had a dog.'* (Feeding Fluffy, from Totem Field Storyboards http://totemfieldstoryboards.org)
nǐ   kěnéng   yǒu   yī-zhī   gǒu.
you   EPIS.POS   have   one-CL   dog
'You might have had a dog.'

[6] (16) and (19) can only be translated with 'will' as the English epistemic necessity modal 'must' can't be future-oriented (Portner 2009: 230).

(21) Context: *You and your friend agreed to meet at 41 St.'s 7-11, but you didn't see him at the appointed time. The 7-11 clerk told you there's another 7-11 on 41 St. so you hastened to go there but still didn't find him. When you came home, you got his call. He says, 'Why didn't you wait for me? I was only 15 minutes late!' You reply:*
nǐ kěnéng qù-*le* lìngwài yī-jiā 7-11.
you EPIS.POS go-PFV another one-CL 7-11
'You might have gone to another 7-11.'

(22) Context: Y*ou thought you were going to meet your friend at 41 St.'s 7-11, but you didn't see him at the appointed time. You didn't have a cell with you so you only waited there but never find him. Later when you came home, you got his call, saying 'Why didn't you go find a booth and call me? I was waiting for you at 44 St.'s 7-11 for 1 hour!' You reply:*
nǐ kěnéng *huì* zhǎo-bú-dào wǒ.
you EPIS.POS FUT find-NEG-out me
'You might not have found me' (if I left the 7-11 and you arrived while I was gone).

A future T.P. is also possible for the epistemic modals. The context in (23) says that according to our knowledge at the speech time, it's impossible for human being to go up to the outer space by elevator, but it will become possible later once the Japanese company develops the technology. The future T.P. requires the future morpheme *jiāng* to precede the epistemic modal, which forms a contrast with the future T.O., which requires *huì* to follow the modal, as in (16), (19), and (22).[7]

(23) Context: *Japanese construction company Obayashi wants to build an elevator to space and transport passengers to a station about a tenth the distance to the moon. The elevator would use super-strong carbon nanotubes in its cables and could be ready as early as 2050, according to Tokyo-based Obayashi.* (http://news.cnet.com, Feb. 23, 2012)
2050-nián *jiāng*/**huì* kěnéng dā diàntī shàng wàitàikōng.
2050-year FUT EPIS.POS take elevator go.up outer.space
'(Human being) will be likely to go up to the outer space by elevator by 2050.'

We have seen that the Mandarin epistemic modals allow a present, past and future T.O. via the aspectual or temporal marking, and can involve a past T.P. without any particular grammatical element and a future T.P. with the future morpheme *jiāng* above the modal. The temporal interpretation possibilities for epistemic modals are summarized in Table 3.

Table 3. T.P. and T.O. for epistemic modals

	T.P.		Temporal Orientation		
	Past/Present	Future	Past	Present	Future
State	---	*jiāng*	*past adverb*	*(covert) imperfective*	*huì*
Event			*perfective*	*imperfective*	

4.2 Circumstantial modals

Circumstantial modals differ from the epistemic modals in being incompatible with the future auxiliary *huì* in a context of future T.O., irrespective of whether their ordering source is deontic (24), teleological (25), or empty (26), and of whether their prejacent is stative or eventive, (24a) vs. (24b).

(24) a. Context: *You plan to go on a short trip in the coming weekend but your mother rejects your plan because this weekend is your grandfather's 60th birthday and everyone should be there.*
nǐ bìxū (**huì*) zài jiā.
you CIRC.NEC FUT LOC home

[7] The morphemes *jiāng* and *huì* are sometimes exchangeable without changing the future meaning, or even coexist as a disyllabic morpheme, as in (i). I leave the difference between them for future research.
(i) tā *jiāng/huì/jiānghuì* chéngwèi yōuxiù de lǎoshī.
he FUT become excellent POSS teacher
'He will become an excellent teacher.'

'You must be at home.'

b. Context: *We are playing a game and agree that the loser will sing a song for our entertainment. It comes out that John is the loser.*
tā bìxū (**huì*) chàng gē.[8]
he CIRC.NEC FUT sing song
'He must sing.'

(25) Context: *Your friend asks you to taste her new recipe and give her advice.*
nǐ kěyǐ (**huì*) jiā duō yīdiǎn yánbā.
you CIRC.POS FUT add more a.little salt
'You can add some more salt.'

(26) Context: *You acquire a piece of land in a far away country and discover that the soil and climate are very much like at home, where hydrangeas prosper everywhere. Since hydrangeas are your favorite plants, you wonder whether they would grow in this place and inquire about it.* (Kratzer 1991: 646)
xiùqiúhuā kěyǐ (**huì*) shēngzhǎng zài zhèlǐ.
hydrangea CIRC.POS FUT grow LOC here
'Hydrangeas can grow here.'

In addition, the circumstantial modals cannot embed a prejacent taking a perfective marker, which is the source of a past T.O. in the epistemic modals, as in (27). What if the circumstantial modal embeds an achievement verb, which is assumed to have a covert perfective aspect in Lin (2006) (cf. Section 3), as in (28)? If there were a covert perfective, it would enforce a past T.O., but the sentence only allows a future T.O.

(27) * tāmen bìxū chàng-*le*/*guò* gē.
they CIRC.NEC sing-PFV/EXP song
'They had to sing songs.'

(28) tāmen bìxū dǎ-yíng bǐsài.
they CIRC.NEC play-win game
'They must win the game.'
#'They had to win the game.'

A past time adverb can't lead to a past T.O. but instead triggers a counterfactual reading. For instance, (29) conveys that at the mentioned time yesterday, in view of their goal, that they win the game is true in all accessible worlds, but this was not realized in the actual word, thus involving past obligation, a past T.P. Therefore, the unavailability of a past T.O. in the circumstantial modals is evidenced by the absence of a perfective aspect with the prejacent as well as the fact that a past time adverb only shifts the temporal perspective.

(29) Context: *In order to reach the final champion match, they had to win the game yesterday (but they lost in the end).*
tāmen *zuótiān* bìxū dǎ-yíng bǐsài.[9]
they yesterday CIRC.NEC play-win game
'They had to win the game yesterday' (but they didn't).

---

[8] *huì* is grammatical here if it's interpreted as an ability modal and the sentence will have multiple modals:
(i) tā bìxū *huì* chàng gē.
he CIRC.NEC ABIL sing song
'He must be able to sing songs.'

[9] The counterfactual reading can be facilitated with a contrastive adverb *běnlái* or a sentence-final particle/mood *de*:
(i) tāmen zuótiān *běnlái* bìxū/kěyǐ dǎ-yíng bǐsài *de*.
they yesterday originally CIRC.NEC/CIRC.POS play-win game PRT
'They had to have won a game yesterday (but they didn't).'

The imperfective markers are the only viewpoint aspects allowed under circumstantial modals. When an imperfective aspect is attached to an eventive prejacent, we see either a future or a present T.O. is available (30), depending on the context. A stative prjacent, which is argued to only take a covert imperfective marker (Lin 2006, cf. (5a) vs. (6a)), can also allows a present T.O. (31), in addition to the future T.O. (24a).

(30) a. Context: *Our music class has been converted into an additional mathematics class when the midterm approaches. However, the school inspector will make his rounds in our music class today. Our teacher asks us:*

nǐ-men děngyīxià bìxū (*huì) zài chàng gē. FUTURE T.O.
you-PL later CIRC.NEC FUT PROG sing song
'You guys must be singing later.'

b. Context: *The school inspector arrives at the school and knows from the students' course schedule that now is their music class. He thinks:*

tāmen xiànzài bìxū *zài* chàng gē. PRESENT T.O.
they now CIRC.NEC PROG sing song
'(According to the educational policy) they must be singing now.'

(31) xiànzài shì wǔxiū, tóngxué-men dōu bìxū zài jiàoshì.
now be lunch.break class-PL all CIRC.NEC LOC classroom
'Now is lunch break so all students have to be in the classroom.'

With regards to T.P., the circumstantial modals are parallel to the epistemic modals. We have seen above that the circumstantial modals allow both a present T.P. (24-26), and (30-31), and a past T.P. (29). They also allow a future T.P. if they are preceded by the future maker *jiāng* (32).

(32) Context: *According to current law, foods containing more than 0.5g of trans fat must say so. But if a snack contains, for example, 0.49 g of trans fat, it can be labeled to indicate that it contains none at all. Considering the daily maximum 1.11 g of trans fat, you could exceed that amount by eating just three servings of "zero trans fat" snack food. The researcher and medical student at Case Western Reserve University School of Medicine is calling for a change in that policy.* (http://www.cbsnews.com, Jan. 4, 2011)

shípǐn suǒ hán fǎnshìzhī *jiāng* bìxū biāozhù.
food REL contain trans.fats FUT CIRC.NEC label
'Food which contains trans fats will have to be labeled.'

The temporal interpretation possibilities for the circumstantial modals are summarized in Table 4. The circumstantial modals have been shown to allow a future and present T.O. but not a past T.O. In this they contrast with the epistemic modals, which allow all the three kinds of T.O, as illustrated by the presence vs. absence of a perfective aspect under the two types of modals. Moreover, unlike the epistemic modals, a future T.O. for the circumstantial modals doesn't permit any overt marker on both stative and eventive prejacents. A stative prejacent can additionally have a present T.O. by itself, while an eventive prejacent can be present-oriented if an imperfective aspect is present.

Table 4. T.P. and T.O. for circumstantial modals

<table>
<tr><th></th><th colspan="2">T.P.</th><th colspan="3">Temporal Orientation</th></tr>
<tr><th></th><th>Past/Present</th><th>Future</th><th>Past</th><th>Present</th><th>Future</th></tr>
<tr><td>State</td><td rowspan="2">---</td><td rowspan="2">jiāng</td><td rowspan="2">*</td><td>(covert) imperfective</td><td rowspan="2">null</td></tr>
<tr><td>Event</td><td>imperfective</td></tr>
</table>

4.3 The ambiguous weak necessity modal

We have shown that the aspectual and future marking under both types of modals exhibit contrastive occurrence restriction: The epistemic modals co-occur with a perfective marker and the future marker *huì*, while the circumstantial modals don't. This generalization will predict that the weak necessity modal *yīnggāi*, which is ambiguous between an epistemic and circumstantial modal base, will be interpreted differently in terms of these markings. The prediction is upheld. When *yīnggāi* co-occurs with a perfective marker (33) or the future marker (34), it can only be epistemic.

(33) Context: *You were watching the Canucks and in the second period, they were up 4-1, but you fell asleep so didn't know the exact result.*
tāmen yīnggāi dǎ-yíng-*le* bǐsài. EPISTEMIC/*DEONTIC: PAST
they W.NEC play-win-PFV game
'They should have won the game.'

(34) Context: *They have been practicing very hard this season; their competitors haven't received intensive training; they are also the host team...*
tāmen yīnggāi *huì* dǎ-yíng bǐsài. EPISTEMIC/*DEONTIC: FUTURE
they W.NEC FUT play-win game
'They should win the game.'

On the contrary, if *yīnggāi* combines with a bare prejacent, it can only be used as circumstantial, with possibly varied ordering sources (35).

(35) tāmen yīnggāi dǎ-yíng bǐsài.
they W.NEC play-win game
'They should win the game.'
DEONTIC: *Their team is the best in the country. Given the convention that the best team always wins the international championship, they should win the game.*
TELEOLOGICAL: *Given their goal of being recognized by people worldwide, they should win the game.*
BOULETIC: *In view of the high value prizes they want, they should win the game.*

As for the imperfective aspects, since they are able to attach to prejacents of both types of modals, yielding a present T.O for the epistemic modals, and a present or future T.O. for the circumstantial modals, we predict that *yīnggāi* can co-occur with an eventive prejacent with the progressive marker *zài* and a stative one with a covert imperfective aspect, and the sentences can be used in types of modal contexts. This is upheld (36-37).

(36) tāmen yīnggāi *zài* chàng gē.
they W.NEC PROG sing song
'According to the schedule, they should be singing.' EPISTEMIC: PRESENT
'According to the regulation, they should be singing.' DEONTIC: PRESENT/FUTURE

(37) āmen yīnggāi zài jiā.
they W.NEC LOC home
'According to the schedule, they should be at home.' EPISTEMIC: PRESENT
'According to their mother's order, they should be at home.' DEONTIC: PRESENT/FUTURE

The weak necessity modal *yīnggāi* thus reinforces the generalizations we have made: The two types of modals differ in whether they allow a past T.O., as evidenced by the presence vs. absence of a perfective aspect under the modals. A future T.O. for the two types of modals is derived differently: via *huì* under the epistemic modals, whereas via the null morpheme under the circumstantial ones. A present T.O. for both types of modals is possible only with a stative verb, and an imperfective aspect. The findings give empirical evidence for the correlation between the types of modal bases, epistemic vs. non-epistemic, and T.O. argued for in the literature (Condoravdi 2002[10], Stowell 2004, Werner 2006, van de Vate 2010,

[10] Condoravdi (2002) actually proposes that the temporal asymmetry is between epistemic and metaphysical modal bases, but Abusch (to appear) argues that the modals which are always future-oriented use a circumstantial but not metaphysical modal base because not all facts about the base world are taken into account. Since Mandarin lexically distinguishes epistemic and circumstantial bases, the temporal asymmetry

Matthewson to appear, among others).

5 A split analysis

What kind of modal semantics can best account for the temporal difference based on the facts discussed so far? Since the future marker *huì* is obligatorily absent under the circumstantial modals, I argue that the future T.O is encoded in the semantics of the circumstantial modals. On the other hand, the varied T.O. of the epistemic modals is determined by the presence of the aspectual/future marking.

There are reasons for not adopting a uniform analysis which either attributes the temporal semantics to both modals or to a separate future/prospective morpheme. The former can be represented by Condoravdi's (2002) analysis, which argues that modals in English are all inherently forward-shifting and different interpretations other than future are conditioned by eventuality and the presence vs. absence of the perfect. This proposal cannot account for the fact that the epistemic modals in Mandarin must rely on the presence of the future marker to have a future T.O. The Mandarin epistemic modals thus differ from the English ones with respect to how to express futurity in a simple modal sentence. However, the Mandarin fact is not very surprising. In Gitksan, a Tsimshian language spoken in north-western British Columbia of Canada, it's also found that a future T.O. for epistemic modals always needs a prospective aspect morpheme *dim* on the prejacent (Matthewson to appear).

Although Mandarin and Gitksan are parallel with the epistemic modals, they are different particularly with respect to the presence of the future marker under circumstantial modals: Gitksan obligatorily requires the prospective aspect marker *dim* whereas Mandarin doesn't permit the future marker *huì*. To account for the obligatory presence of a prospective aspect marker under all modals with a future T.O. in Gitksan, Matthewson (to appear) argues that Gitksan modals have no inherent future orientation, and it's *dim* which appears under a modal that gives a future orientation. Applying a similar analysis to the Mandarin modals would require the future marker *huì* to be responsible for a future T.O. for both types of modals. However, this wrongly predicts that the future marker is obligatorily present under the circumstantial modals. Crucially, the split condition of the future marker in each modal class suggests the future semantics is lexically encoded only in circumstantial modals (Enç 1996, van de Vate 2010, Abusch to appear, among others).

Due to the difficulties of applying a uniform analysis, I suggest that Condoravdi's (2002) analysis of English modals can only be extended to the Mandarin circumstantial modals, while Matthewson's (to appear) analysis of Gitksan modals can only be applied to the Mandarin epistemic modals. Adopting a split analysis, I propose that in Mandarin, circumstantial modals encode futurity in their semantics, whereas epistemic modals only introduce a time variable but don't specify its reference, which is derived from semantic composition with the viewpoint aspects and the future marker *huì*. Following Rullmann et al. (2008), the modals are defined as lexically restricted to either circumstantial or epistemic modal bases from the very first in the lexicon. I start with the analysis of the circumstantial modals.

Lexical entries for the Mandarin circumstantial necessity and possibility modals are given in (38-39). According to Abusch (1998), [t, _) designates an interval with t as an initial subinterval and extending to the end of time, which gives forward-shifting semantics. Condoravdi's AT relation (40), how property P is instantiated in world w at time t, is also assumed here because in Mandarin, the eventuality type makes a difference with respect to whether the temporal orientation can be present or not.

(38) [[$bìxū_{MB}$]] is defined only if MB is circumstantial. If defined,
[[$bìxū_{MB}$]] = $\lambda P_{<i, st>}\ \lambda w\ \lambda t\ \forall w'\ [w' \in MB(w, t) \rightarrow AT([t, \_), w', P)]$

(39) [[$kěyǐ_{MB}$]] is defined only if MB is circumstantial. If defined,
[[$kěyǐ_{MB}$]] = $\lambda P_{<i, st>}\ \lambda w\ \lambda t\ \exists w'\ [w' \in MB(w, t)\ \&\ AT([t, \_), w', P)]$ (Condoravdi 2002:71)

(40) $$AT(t, w, P) = \begin{cases} \exists e\ [P(w)(e)\ \&\ \tau(e, w) \subseteq t] & \text{if P is eventive} \\ \exists e\ [P(w)(e)\ \&\ \tau(e,w)\ o\ t] & \text{if P is stative} \\ P(w)(t) & \text{if P is temporal} \end{cases}$$ (Condoravdi 2002:70)

between the two provides evidence for Abusch's (to appear) argument.

(41-42) are the results of applying the analysis to the deontic necessity modal with a stative and eventive prejacent respectively: (41) states that for all accessible worlds compatible with what the rules require at the time t, the time of the stative prejacent 'you are at home' overlaps the interval between the time t and the end of time. The overlapping relation gives either a present T.O. if the state starts at some point in the past of the time t, or a future T.O. if it is totally included in the interval [t, _). (42) shows that the time of the eventive prejacent is fully included in the interval [t, _) due to the semantics of AT, thus yielding a future T.O. alone. The time variable t can be the speech time or a past topic time if the sentence has a past T.P.

(41) a. nǐ bìxū zài jiā. PRESENT/FUTURE T.O.
you CIRC.NEC LOC home
'You must be at home.'

b. [[*nǐ bìxū*$_{MB}$ *zài jiā*]]
$= \lambda t\ \lambda w\ \forall w'\ [w' \in MB(w, t) \rightarrow \exists e\ [[\text{you are at home}](w', e)\ \&\ \boldsymbol{\tau(e, w')\ o\ [t, \_)}]]$

(42) a. tā bìxū chàng gē. FUTURE T.O.
he CIRC.NEC sing song
'He must sing songs.'

b. [[*tā bìxū*$_{MB}$ *chàng gē*]]
$= \lambda t\ \lambda w\ \forall w'\ [w' \in MB(w, t) \rightarrow \exists e\ [[\text{he sings songs}](w', e)\ \&\ \boldsymbol{\tau(e, w') \subseteq [t, \_)}]]$

Note that an eventive prejacent only acquires a present T.O. if it takes an imperfective aspect (43) (repeated from (30) above):

(43) nǐmen bìxū *zài* chàng gē. PRESENT/FUTURE T.O.
you CIRC.NEC PROG sing song
'You guys must be singing.'

The effect of eventuality is also observed in English epistemic modals, where a progressive prejacent behaves like a stative one with respect to allowing a present T.O. I assume the temporal property of the imperfective aspect and an eventive predicate follows the third sub-relation of AT in (40). Applying Lin's (2006) semantics of the imperfective aspect (44) (repeated from (7a) above),[11] and the lexical entry of the modal (38), the computation of (43) is shown in (45). The time t can be instantiated by a present or past T.P. and the relation that the event time t' includes the interval [t, _) gives a present or future T.O.

(44) $[[\text{IMPRF}]] = \lambda P\langle i, st\rangle\ \lambda t\ \lambda w\ \exists t'\ [t \subseteq t'\ \&\ \exists e\ [P(w)(e)\ \&\ \tau(e, w) = t']]$

(45) a. nǐmen bìxū *zài* chàng gē. PRESENT/FUTURE T.O.
you CIRC.NEC PROG sing song
'You guys must be singing songs.'

b. [[*nǐmen bìxū*$_{MB}$ *zài chàng gē*]]
$= \lambda t\ \lambda w\ \forall w'[w' \in MB(w, t) \rightarrow \exists t'\ [\boldsymbol{[t,\_) \subseteq t'}\ \&\ \exists e[[\text{you guys sing songs}](w', e)\ \&\ \tau(e, w') = t']]$

Next, lexical entries for the epistemic necessity and possibility modals can be spelled out in (46-47). Unlike the circumstantial modals, no AT relation and specified interval [t, _) is encoded in the epistemic modals.

(46) [[*yīdìng*$_{MB}$]] is defined only if MB is epistemic. If defined,
[[*yīdìng*$_{MB}$]] $= \lambda P_{<i, st>}\ \lambda w\ \lambda t\ \forall w'\ [w' \in MB(w, t) \rightarrow P(t)(w') = 1]$

[11] Since the semantics of imperfective/perfective viewpoint aspect and the future morpheme *huì* given in Lin (2006) (cf. (7) and (11) above) doesn't have a world and event variable, I convert them into (44), (49), and (51).

(47) [[*kěnéng*$_{MB}$]] is defined only if MB is epistemic. If defined,
[[*kěnéng*$_{MB}$]] = $\lambda P_{<i, st>}\ \lambda w\ \lambda t\ \exists w'\ [w' \in MB(w, t)\ \&\ P(t)(w') = 1]$ (Matthewson to appear: 7)

Now I demonstrate the analysis applying to the epistemic possibility modal with a present T.P. (48) is repeated from (17) above, in which the semantics is composed of an epistemic possibility modal base (47) and the imperfective aspect *zài* (44). (48) asserts that there's a world w' which is accessible from w at the speech time, in which 'they drink wine' which includes the speech time, thus giving a present T.O.

(48) a. tāmen kěnéng *zài* hē jiǔ. PRESENT T.O.
they EPIS.POS PROG drink wine
'They may be drinking wine.'

b. [[t*āmen kěnéng*MB *zài hē jiǔ*]]
$= \lambda w\ \exists w'\ [w' \in MB(w, now) \&\ \exists t'\ [\mathbf{now \subseteq t'} \&\ [\exists e\ [[\text{they drink wine}](w')(e) \&\ \tau(e, w') = t']]]]$

When the epistemic possibility modal embeds a predicate taking the perfective aspect (49) (repeated from (7b) above), it asserts that there's a world w′ which is accessible from w at the speech time, in which 'they drink wine' at a time t″ which is included in a past topic time t (50). The precedence relation t < now combing with the inclusion relation t″ ⊆ t gives the past T.O. and such relations exclusively arise from the semantics of the perfective aspect.

(49) $[[\text{ PRF}]] = \lambda P_{<i, st>}\ \lambda t\ \lambda t'\ \lambda w\ \exists t''\ [t'' \subseteq t\ \&\ \exists e\ [P(w)(e)\ \&\ \tau(e, w) = t''] \&\ t < t']$

(50) a. tāmen kěnéng hē-*le* jiǔ. PAST T.O.
they EPIS.POS drink-PFV wine
'They may have drunk wine.'

b. [[*tāmen kěnéng*$_{MB}$ *hē le jiǔ*]]
$= [[\textit{kěnéng}_{MB}]](\lambda t'\ \lambda w\ \exists t\ \exists t''\ [t'' \subseteq t\ \&\ \exists e\ [P(w)(e)\ \&\ \tau(e, w) = t'']\ \&\ t < t'])$[12]
$= \lambda w\ \exists w'\ [w' \in MB(w, now)\ \& \exists t\ \exists t''\ [\mathbf{t'' \subseteq t}\ \&\ [\exists e\ [[\text{they drink wine}](w')(e)\ \&\ \tau(e, w') = t'']\ \&$
$\mathbf{t < now}]]]$

A future T.O. with the epistemic modal is derived from the future marker *huì* which encodes a precedence relation between the speech time and event time. The semantics of *huì* is repeated in (51) from (11) above. Applying (47) and (51), (52) asserts that there's a world w' which is accessible from w at the speech time, such that 'they drink wine' at a time t' which follows the speech time.

(51) [[*hui*]] $= \lambda P_{<i, st>}\ \lambda t\ \lambda t'\ \lambda w\ [\exists e\ [P(w)(e)\ \&\ \tau(e, w) = t]\ \&\ t' < t]$

(52) a. tāmen kěnéng *huì* hē jiǔ. FUTURE T.O.
they EPIS.POS FUT drink wine
'They may drink wine.'

b. [[*tāmen kěnéng*$_{MB}$ *huì hē jiǔ*]]
$= [[\textit{kěnéng}_{MB}]](\lambda t'\ \lambda w\ \exists t\ [\exists e\ [P(w)(e)\ \&\ \tau(e, w) = t]\ \&\ t' < t])$
$= \lambda w\ \exists w'\ [w' \in MB(w, now)\ \&\ \exists t\ [\exists e\ [[\text{they drink wine}](w')(e)\ \&\ \tau(e, w') = t]\ \&\ \mathbf{now <} t]]$

In contrary to the epistemic modals, the incompatible co-occurrence of a circumstantial modal and the future marker is explained by the inherent semantics of the circumstantial modal. As demonstrated in (53) (repeated from (24b)), since the circumstantial modal already encodes in its lexical entry an interval [t′, _), the composition of the modal and the future marker *huì*, which denotes the precedence of a topic time

[12] Lin (2006) assumes a rule applies at the IP level to an output translation of type <i, <i, t>>, closing an unfilled topic time variable, because "IP is the level where a topic time should be found" (Lin 2006: 5). Here I assume that such a rule in modal sentences applies at the ModP level, as in both (50) and (53), because T.O. is parallel to the semantics of the perfective aspect/future marker, cf. Section 4.1.

over the evaluation time t′ < t, gives [t′, _) < t and causes a conflict: a topic time can't precede an interval extended to the end of time.

(53) a. *tā bìxū *huì* chàng gē.
he CIRC.NEC FUT sing song
'He must sing songs.'

b. [[*tā bìxū*$_{MB}$ *huì chàng gē*]]
= [[*bìxū*$_{MB}$]](λt′ λw ∃t [∃e [[he sings songs]](w)(e) & τ(e, w) = t] & t′ < t])
= λt′ λw ∀w′ [w′ ∈ MB(w, t′)→ ∃t [∃e [[they drink wine]](w′)(e) & τ(e, w′) = t] & **[t′, _) < t**]
= λw ∀w′ [w′ ∈ MB(w, now) → ∃t [∃e [[they drink wine]](w′)(e) & τ(e, w′) = t] & **[now, _) < t**]

The split analysis thus accounts for the correct temporal interpretations of Mandarin modals. It captures the distinct occurrence of the future marker under each modal.[13] To recapitulate, the interval [t′, _) encoded in the lexical entry of circumstantial modals makes circumstantial modals incompatible with attaching to the future marker. The lack of such an interval in epistemic modals explains why epistemic modals lack this incompatiblity.

6 Conclusion and remaining issues

The temporal orientation of modals in Mandarin relates to the types of modal bases: Circumstantial modals allow a present or future orientation, while epistemic modals allow a past, present and future T.O. The hypothesis is proved by the co-occurrence restriction between the perfective/future markers and the modals, and correctly predicts uses of the ambiguous weak necessity modal. Past time adverbs also exhibit a contrast between the two types of modals: while they are able to bring about a past T.O. for epistemic modals, they agree with a past T.P. for circumstantial modals. I propose that the obligatory absence and presence of the future marker *huì* under the circumstantial and epistemic modals in Mandarin need a split analysis of future semantics for each modal.

A remaining question is why the languages vary in the ways of expressing the future tendency for circumstantial modals. Specifically, Mandarin and Gitksan differ from English in the presence of a separate future morpheme in epistemic modals, while yet Mandarin deviates from Gitksan in the absence of a future morpheme in circumstantial modals, as summarized in Table 5.

Table 5. A comparison of futurity for the English, Gitksan, and Mandarin modals

	Epistemic: Future T.O.	Circumstantial
English	null	
Gitksan	*dim*	*dim*
Mandarin	*huì*	*null*

I suggest that the variations might relate to the syntactic properties of modals in individual languages. In Mandarin, multiple modals exhibit an ordering restriction, which indicates a hierarchy of modals in the syntactic structure; particularly, epistemic modals precede deontic and ability modals (Lin 2012). The hierarchical difference could explain why only epistemic modals, which are projected higher

[13] This analysis can't explain why the perfective aspect is incompatible with circumstantial modals (Matthewson, p.c.); actually, it instead predicts that the perfective aspect can appear under a circumstantial modal (i) (repeated from (27)):
(i) a. *tāmen bìxū chàng-*le* gē.
they CIRC.NEC sing-PFV song
'They had to sing songs.'
b. = [[*tāmen bìxū*$_{MB}$ *chàng-le gē*]]
= [[*bìxū*$_{MB}$]](λt′ λw ∃ t ∃ t′′ [t′′ ⊆ t & ∃e [[they sing songs]](w)(e) & τ(e, w) = t′′] & t < t′])
= λt′ λw ∀w′ [w′ ∈ MB(w, t′) →∃t ∃t′′ [t′′ ⊆ t & ∃e [[they sing songs]](w′)(e) & τ(e, w′) = t′′] & **t < [t′, _)**]]
An alternative is to adopt Matthewson's (to appear) proposal that the futurity of modals comes from a separate prospective aspect marker, which in Mandarin is covert. This would ideally explain why the perfective aspect is incompatible with the circumstantial modals in Mandarin, as they already have a viewpoint aspect. However, the challenge to the prospective aspect analysis is to explain why Mandarin possess both a null prospective with circumstantial modals and an overt one with epistemic modals.

than circumstantial modals, can co-occur with the perfective aspect and future *huì*. A crucial support comes from a parallel property between the circumstantial modal and *huì*. Lin (2011) argues that epistemic modals in Mandarin take a finite TP complement, and circumstantial modals take a nonfinite TP complement; Lin also shows that the future modal/auxiliary *huì* patterns with circumstantial modals in taking a nonfinite TP complement. If they are the same syntactic head, they are not able to co-occur. Multiple-modal constructions in Mandarin and a specific investigation of the future morpheme *huì* will have a direct bearing on the (in)compatibility of the types of modals and aspectual/future marking.

References

Abusch, Dorit. 1998. Generalizing tense semantics for future contexts. In S. Rothstein (ed.), *Events and Grammar*, pp. 13-33. Dordrecht: Kluwer.

Abusch, Dorit. To appear. Circumstantial and temporal dependence in counterfactual modals. Natural Language Semantics.

Bohnemeyer, J. and M. Swift. 2004. Event realization and default aspect. *Linguistics and Philosophy* 27: 263-296.

Borgonovo, Claudia and Sarah Cummins. 2007. Tensed modals. In Luis Eguren et al. (eds.), *Coreference, Modality, and Focus: Studies on the Syntax-Semantics Interface*, pp. 1-18. John Benjamins.

Chao, Y-R. 1968. *A grammar of spoken Chinese*. University of California Press.

Condoravdi, Cleo. 2002. Temporal interpretation of modals: Modals for the present and for the past. David Beaver et al. (eds.), *The Construction of Meaning*, pp. 59-88. CSLI Publications.

Demirdache, Hamida, and Myriam Uribe-Etxebarria. 2008. On the temporal syntax of non-root modals, In Jacqueline Lecarme et al. (eds.), *Time and Modality*, pp. 79-114. Dordrecht: Springer.

Eide, Kristin M. 2003. Modals and tense. In M. Weisgerber (ed.), *Proceedings of Sinn und Bedeutung* 7, pp. 120-135. The University of Konstanz, Germany.

Enç, Mürvet. 1996. Tense and modality. In S. Lappin (ed.), *The Handbook of Contemporary Semantic Theory*, pp. 345-358. Oxford: Blackwell.

Kratzer, Angelika. 1991. Modality. In Dieter Wunderlich et al. (eds.), *Semantics: An International Handbook of Contemporary Research*, pp. 639-650. Berlin: de Gruyter.

Hacquard, Valentine 2009. On the interaction of aspect and modal auxiliaries. *Linguistics and Philosophy* 32: 279-315.

Hu, Jianhua, Pan, Haihua, and Xu, Liejiong. 2001. Is there a finite-nonfinite distinction in Chinese? *Linguistics* 39: 1117-1148.

Laca, Brenda. 2008. On modal tenses and tensed modals. In Chiyo Nishida et al. (eds.), *Proceedings of CHRONOS 2008*.

Li, Charles N. and Thompson, Sandra A. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Li, Ren-Zhi. 2004. *Modality in English and Chinese: A typological perspective*. Doctoral Dissertation, University of Antwerp.

Lin, T.-H. 2011. Finiteness of clauses and raising of arguments in Mandarin Chinese. *Syntax* 14: 48-73.

Lin, T.-H. 2012. Multiple-modal constructions in Mandarin Chinese and their finiteness properties. *Journal of Linguistics*: 1-36.

Lin, J. W. and J. C. C. Tang. 1995. Modals as verbs in Chinese: A GB perspective. *Collection of Sinica Academia* 66: 53-105. (in Chinese)

Lin, Jo-wang. 2003. Temporal reference in Mandarin Chinese. *Journal of East Asian Linguistics* 12: 259-311.

Lin Jo-wang. 2006. Time in a language without tense: The case of Chinese. *Journal of Semantics* 23:1-56.

Matthewson, Lisa and Hotze Rullmann. 2012. Epistemic modals can scope under past tense. Paper presented at the Workshop on Modality, April 20-21, Ottawa, Canada.

Matthewson, Lisa. 2004. Methodology of semantic fieldwork. *International Journal of American Linguistics 70*: 369-415.

Matthewson, Lisa. To appear. On the (non-)future orientation of modals. In A. Aguilar-Guevara et al. (eds.), *Proceedings of Sinn und Bedeutung 16*. Cambridge, MA: MIT Working Papers in Linguistics.

Portner, Paul. 2009. *Modality*. Oxford University Press.

Ren, F. 2008. *Futurity in Mandarin Chinese*. Doctoral dissertation, University of Texas at Austin.

Rullmann, Hotze, Lisa Matthewson, and Henry Davis. 2008. Modals as distributive indefinites.

Natural Language Semantics 16: 317-357.
Smith, S. Carlota, and Erbaugh Mary S. 2005. Temporal interpretation in Mandarin Chinese. *Linguistics* 42: 713-756.
Stowell, Tim. 2004. Tense and modals. In Jacqueline Gueron et al. (eds.), *The Syntax of Time*, pp. 621-636. Cambridge: MIT Press.
Sybesma, Rint. 2007. Whether we tense-agree overtly or not. *Linguistic Inquiry* 38: 580-587.
Tsang, Chui-Lim. 1981. *A semantic study of modal auxiliary verbs in Chinese*. Doctoral dissertation: Stanford University.
Vander Klok, Jozina. 2008. Javanese modals: In between auxiliaries and verbs. Paper presented at the annual conference of the Canadian Linguistics Association, May 31-June 2, UBC, Vancouver.
van de Vate, Marleen. 2010. Possibility Modality in Saamaka. In Lima, Suzi (ed.), *UMOP 41: Proceedings of SULA V*.
Werner, Tom 2006. Future and non-future modal sentences. *Natural Language Semantics* 14: 35-255.

Sihwei Chen
sw.chen@alumni.ubc.ca

Learning consonant harmony in artificial languages: Locality[1]

Kevin McMullin
UBC Department of Linguistics

This study uses an artificial language learning experiment to investigate consonant harmony from a learnability perspective. With respect to locality, and the distance between two interacting consonants, the typology of languages with consonant harmony includes two types of systems. First are those that apply harmony in local cases only (i.e. separated by at most one vowel). Second, there are languages in which harmony applies anywhere within some domain, irrespective of distance. Monolingual English speakers are able to learn both types of patterns in the lab, and furthermore, they generalize from limited exposure in ways that mirror the typology. This is taken as evidence that humans have certain learning biases that not only affect how they learn and generalize patterns, but that these biases, over time, can shape the typology of phonological patterns.

1 Introduction

1.1 Phonological patterns and learnability

A great deal of research in linguistics concentrates on identifying, describing, and analyzing phonological patterns. However, we still do not know how humans learn them. This paper focuses on the learning of one such pattern, consonant harmony, in which two consonants in a word are required to agree in some way. For example, Yaka, a language from central Africa, has a perfective suffix *-ili*, which attaches to a verb. However, if the verb contains a nasal consonant, then the [l] in the suffix must become a nasal [n]. Thus, the verb stem *ja**n**-a* 'cry out in pain' is *ja**n**-i**n**i* in the perfective rather than **ja**n**-i**l**i* (Hansson 2010; Hyman 1995). More than 130 languages are known to have some form of consonant harmony system, each with its own set of properties. Some of these properties are more common than others, but we do not know why. One possible explanation is that it is an issue of learnability; the rare patterns are simply harder to learn, so they are less likely to ever arise in a language, let alone persist over time. Current research in linguistics and cognitive psychology has lent support to this idea by showing that some patterns involving the interaction of non-adjacent sounds are indeed more difficult to learn than others (Creel *et al*. 2004; Newport and Aslin 2004), particularly with respect to the relative similarity between the sounds (Gebhart *et al*. 2009; Moreton 2012). This generalization is mirrored in the typology of consonant harmony, as two similar consonants like [s] and [ʃ], or [l] and [r] are much more likely to interact than two dissimilar consonants such as [m] and [k], or [s] and [b]. The present study goes beyond the issue of similarity and investigates the learnability of consonant harmony with respect to locality, hypothesizing that humans will learn these patterns and generalize them in a way that matches their distribution across the world's languages.

1.2 The typology of locality in consonant harmony

This section provides a brief description of the typology of sibilant harmony patterns (see Hansson 2001, 2010a; Rose and Walker 2004 for comprehensive studies of the typology of consonant harmony), focusing especially on generalizations that can be made about locality (i.e. the relative distance between two interacting consonants). Languages that have a consonant harmony system require two interacting consonants to agree with respect to some feature. The relevant feature varies by language and can range from voicing, to nasality, to palatalization, and so on. The most common type of consonant

[1] The research reported here was supported by Standard Research Grant 410-2009-2831 from the Social Sciences and Humanities Research Council of Canada, awarded to Gunnar Ólafur Hansson.

harmony, which is the one used in this study, is sibilant harmony, which affects sibilant consonants such as {s,ʃ,z,ʒ,ts,dz,tʃ,dʒ}. Languages with sibilant harmony often prohibit the co-occurrence of a sibilant from each of the following sets {s,z,ts,dz} and {ʃ,ʒ,tʃ,dʒ}. This can be a purely phonotactic constraint, such that no word would ever have a possibility of violating it, or it can be seen in morphophonological alternations.

When considering the issue of locality, it seems plausible that languages would be more likely to apply consonant harmony to two consonants that are relatively close together in a word, because the pattern should be easier for the learner to detect. With this line of thinking, consonant harmony patterns that apply over longer distances would become increasingly more rare. Interestingly, however, this is only partially the case. Languages apply harmony either to local consonant pairs only (separated by at most one vowel) or to all cases within a word, both local and non-local, no matter the distance between them. This is an important typological split from a learnability perspective, as the range of possible consonant harmony patterns seems to be restricted by something other than purely computational limitations. One could easily imagine a language that has a dependency that holds between two consonants when they are separated by two vowels, but not by three vowels. No such system is known to exist.

Any theory of phonology that we have should account for why languages make this distinction between local and non-local dependencies, but nothing in between, as well as why no language has a non-local dependency that does not hold at local distances. In section 4, I will argue that this typological asymmetry is the result of a learning bias that allows humans to discover only these two types of patterns. A summary and evaluation of a formal learning algorithm that can potentially account for this (a precedence learner from Heinz 2010) will also be discussed.

1.3 Artificial language learning experiments

The traditional way to study any linguistic issues related to learning is to observe infants as they acquire the patterns in their native language. However, as many of these patterns are relatively rare, access to children learning the language is not always feasible. As a result, it is becoming increasingly common for researchers to construct artificial languages for experimental studies (Finley and Badecker 2009; Moreton 2012; Moreton and Pater 2011; Nevins 2010). In a typical experiment, subjects complete a training phase in which they are exposed words from the language, followed by a testing phase. Both of these phases can take many different forms. In the testing phase, for example, some experiments have a 2 alternative forced choice task in which the subject must decide which of two words is "correct" or "sounds better". Alternatively, subjects may be asked to produce certain words, and their responses as well as the errors they make can be taken as evidence for what they have or have not learned (e.g. Rose and King 2007).

This methodology creates an accessible way to study any type of phonological pattern, whether it is relatively frequent, relatively rare, or completely unattested across the world's languages. An additional benefit of this method is that it allows the researcher to have control over the input that the learner gets. These are the obvious advantages for those interested in questions about how humans learn, but the methodology is not without criticisms (see Moreton and Pater 2011 for an overview of the findings and criticisms of these experiments). For example, we may not know what biases the learners are bringing in from their own native language, or their language experience as a whole. Furthermore, we still do not know about the relationship between how children and adults learn the patterns in a new language. These are ongoing, unresolved debates that deserve a full investigation of their own, but in the meantime, having a well-constructed control condition can help us to understand what biases a learner might come in with, so that we can build them into any statistical models.

1.4 Previous studies

Of particular interest here are studies that have investigated the learning of consonant harmony at different levels of locality. Finley (2011) examined how adults learn local, transvocalic dependencies in an artificial language with sibilant harmony, as compared with nonlocal dependencies (across two vowels). In the training phase of experiment 1, subjects heard pairs of words. The first word was a two syllable stem of the form cvSv, where S is a sibilant [s] or [ʃ]. The second was the same stem with a suffix -Su, where the sibilant in the suffix was identical to the sibilant in the stem. Thus, all suffixed training items showed evidence of "first-order" harmony. In the testing phase, subjects chose which of two suffixed forms was correct, one of which had matching sibilants (harmony) while the other had mismatched sibilants. Experiment 2 was similar, but learners were trained only on a "second-order" pattern, in which the stems were of the form Svcv. The results indicated an asymmetry that reflects the typology described above. That is, subjects in experiment 1 learned the first-order pattern for both familiar and unfamiliar first-order test items, and did not generalize this to the second-order contexts, in which the sibilants were further apart. Subjects in experiment 2 were successful in learning the second-order pattern and they generalized the pattern inward to the first-order context. The author used these results as evidence for first-order patterns having a "privileged status".

Using methods similar to those of Finley (2011), Finley (2012) expanded her study by presenting the subjects with longer stems to further investigate the role of distance between two interacting consonants. Subjects were trained on pairs of words consisting of a three syllable stem with one sibilant followed by the suffixed form, in which the sibilant in the suffix -Su displayed harmony with the stem. This study makes two important conclusions. First, that learners are able to learn this type of sibilant harmony pattern even when the sibilants are up to five segments away (i.e. Svcvcv-Su); the further away the dependency, the harder it is to learn. Second, when learners are trained on a "second-order" pattern, with cvSvcv-Su words, they generalize to even longer distances: Svcvcv-Su. Based on these results, Finley argues that when learners are tasked with discovering a long-distance interaction, they do so without reference to intervening distance, and so they apply the pattern to all contexts. Crucially, this generalization excludes the local-only patterns that apply harmony when the distance is at most transvocalic.

1.5 The present study

This paper presents the results of another artificial language learning experiment that replicates the findings of Finley (2011; 2012), while some key methodological differences, described in section 2.5, allow for a further investigation of unresolved issues.

2 Methodology

2.1 Subjects

Thirty-six adults aged 18-40 participated in the study (25 female, 8 male, 3 unspecified; mean age of 24). All were native speakers of North American English, reported no speech or hearing disorders and had no experience with a language containing a harmony system. Subjects were compensated $10 for participating in the experiment, which took about 45 minutes to complete.

2.2 Stimuli

Stimuli were recorded by a phonetically trained, male native English speaker. While he knew that the stimuli would be used for an artificial language experiment, he was unaware of the exact topic of study and the hypothesis. Most of the stimuli consisted of three-syllable "verbs" which took the form cvcvcv. There were also six two-syllable verbs of the form cvcv for the practice phase, as described below. Additionally, the same speaker recorded two corresponding suffixed verbs for each root, one of each -su and -ʃi. Consonants in the stems were chosen from sibilants [s, ʃ] as well as stop consonants [p,t,k,b,d,g] and vowels were chosen from [i,e,a,o,u]. The stimuli set was carefully constructed such that there was no other predictable pattern in the data. Crucially, the same number of each vowel and consonant appeared in each position for the verb roots containing [s] or [ʃ].

2.3 Design and procedure

Subjects were told that their task was to practice pronouncing words from a new language and to learn how to say verbs from the language in the past and future tense. Each subject worked through three phases of the experiment.

In the **practice phase**, subjects were told how to conjugate verbs in the language. They then heard, over a set of headphones, six pairs of words for the past tense (a verb root followed by a suffixed form with -su), and six word pairs for the future tense (this time with the suffix -ʃi). The six verb roots in this portion were just two syllables and contained no sibilants. As a result, there was no input with any evidence of an alternation in the practice phase.

In the **training phase**, subjects heard triplets of words, the first of which was always a three-syllable verb root. Half of these contained one sibilant (one quarter with [s], one quarter with [ʃ]) and the other half contained only stop consonants. This was done in part to test how learners perform without 100% of the input containing meaningful information, and in part to allow the other half of the input to be manipulated in future studies. The two subsequent words in each triplet were the four-syllable suffixed forms. Since both suffixes begin with a sibilant, if the verb root also contained a sibilant, one of the suffixed forms would display consonant harmony triggered by the suffix. For example, given the verb *bugaso*, the suffixed forms would be *bugaso-su* and *bugaʃo-ʃi*. Each triplet was presented twice, with the suffixed forms counterbalanced for order and the order of triplets was randomized for each subject. The subjects were asked to repeat each word into a head-mounted microphone that recorded their productions. This was done to allow for possible further analysis, as well as to reinforce the training process. In total, each subject heard and repeated 120 triplets twice each for a total of 720 productions. The words in both the practice phase and the training phase were presented over the headphones only, and the subjects got no information about any semantic content of the words, except that they were verbs that can be conjugated in these two tenses. Subjects were assigned to one of three training conditions, described below in section 2.4, that differed in the types of words presented in their training phase.

In the **testing phase**, subjects heard 30 new verbs, and then completed a forced choice task, in which they were asked to choose the correct conjugated form of the verb. Testing items included 30 suffixed harmonic/non-harmonic verb pairs of the form cvcvcv-Sv. These pairs consisted of ten each with the sibilant in the three different consonantal positions in the root. There were an equal number of -su and -ʃi forms, and the order of presentation was randomized for each subject. It is important to note that all testing items required the subject to choose an alternation in the root in order to identify the suffixed form with harmony (i.e. if the testing root contained a [s], then the corresponding suffixed forms would both have the suffix -ʃi). It would be ideal to have many filler items in the testing phase, as well as harmonic responses that do not require an alternation. However, a confound of that approach is that any items presented in the testing phase effectively become a part of the training, so it is necessary to both limit the testing items and maximize the amount of relevant data from each subject. As a result, subjects were tested only on cases that required the choice of an unfaithful root alternation in favour of harmony with the sibilant in the suffix. Subjects were given 3 seconds after the onset of the final word in the triplet to register a response, and were presented with their response time.

2.4 Training conditions

The subjects were divided into three conditions: *Local*, *Nonlocal*, and *Control*. Stimuli for these three groups differed only in the training phase, with the testing items being the same for all groups. For the portion of the training items that contained sibilants, the *Local* group was trained on verb roots that contained sibilants only in the last consonant position (i.e. cvcvSv), while the *Nonlocal* group was trained on cvSvcv roots. The *Control* group was trained only on verbs that contained no sibilants. This was done to reveal any potential pre-existing biases that English speakers may have toward choosing harmony, or any biases introduced by the experimental procedure.

2.5 Some important differences

The experimental design outlined above contains some unique methodological features that will be pointed out in this section. First, in the conditions where subjects are expected to learn a consonant harmony pattern (the local and nonlocal conditions) only half of the triplets in the training phase contain any evidence of an alternation. The other half can be thought of as filler items, which give no information about sibilant harmony to the learners. This could not, for example, have been done in Finley's studies (described in section 1.4), since having roots with no sibilants would require the choice of a default suffix. The default option would appear more frequently in training than the alternative, and potentially bias the subjects towards the default suffix in the testing phase. The design of the present experiment avoids this problem by having two completely different suffixes (one for past tense and one for future tense), each with a different sibilant, any root that did not contain a sibilant would still have two suffixed forms, one with each sibilant.

Second, the control condition for this study is quite different than in most artificial language learning experiments. If the study is of a purely phonotactic nature, then subjects in the control condition often skip the training phase completely and proceed directly to the testing phase to determine whether they have a preference for or against words adhering to the phonotactic pattern of interest. For studies that present a morphophonological alternation, subjects in the control condition are usually exposed only to the roots, and then are tested on whether they have a preference for one affixed form or another. One confound of these approaches is that the subjects in the control condition have not actually completed the same procedure as those in other conditions. In the present study, all subjects, whether in the control condition or not, are first trained on 720 items before proceeding to the testing phase. This was done in light of the fact that hearing and repeating hundreds of nonsense words can often be quite boring. Having a control condition that has gone through the same procedure allows for a more fair comparison across groups. An additional advantage to this design is that if the subjects in the control condition come in with no biases towards sibilant harmony (and no propensity to make mistakes), then they are expected to never pick a suffixed form that is unfaithful to the root. Rather, the training phase should have taught them simply to add a suffix, and otherwise remain faithful. This is a departure from many other artificial language learning experiments where subjects in the control condition are expected to perform near chance, instead of near 0%.

3 Results and analysis

As reported in section 2.1, thirty-six subjects participated in this study. However, the data from six subjects was dropped (4 Control, 1 Local, 1 Nonlocal) as the result of one of the following: failure to follow instructions (2 subjects), choosing either the first or second test item on all recorded trials (2 subjects), equipment failure (2 subjects). Of the remaining 30 subjects, 827 of a possible 900 responses were registered within the allotted three-second timeframe.

3.1 An overview of responses

The first step in analyzing the results is to get a clear picture of what subjects' choices were in each of the training conditions and for each type of testing item. To do this, I will present the results in this section as the proportion of testing items in which subjects chose to apply consonant harmony rather than stay faithful to the root. Keeping in mind that subjects in the control condition saw no evidence of an alternation in their input, their scores should be close to 0% if they are in no way biased towards harmony. Figure 1 below presents the proportion of responses where the subject chose a non-faithful alternation in the root in order to have harmony with the sibilant in the suffix. The results here are intended to give an overall picture of what the subjects in each condition chose, but will not be used to test for statistical significance.

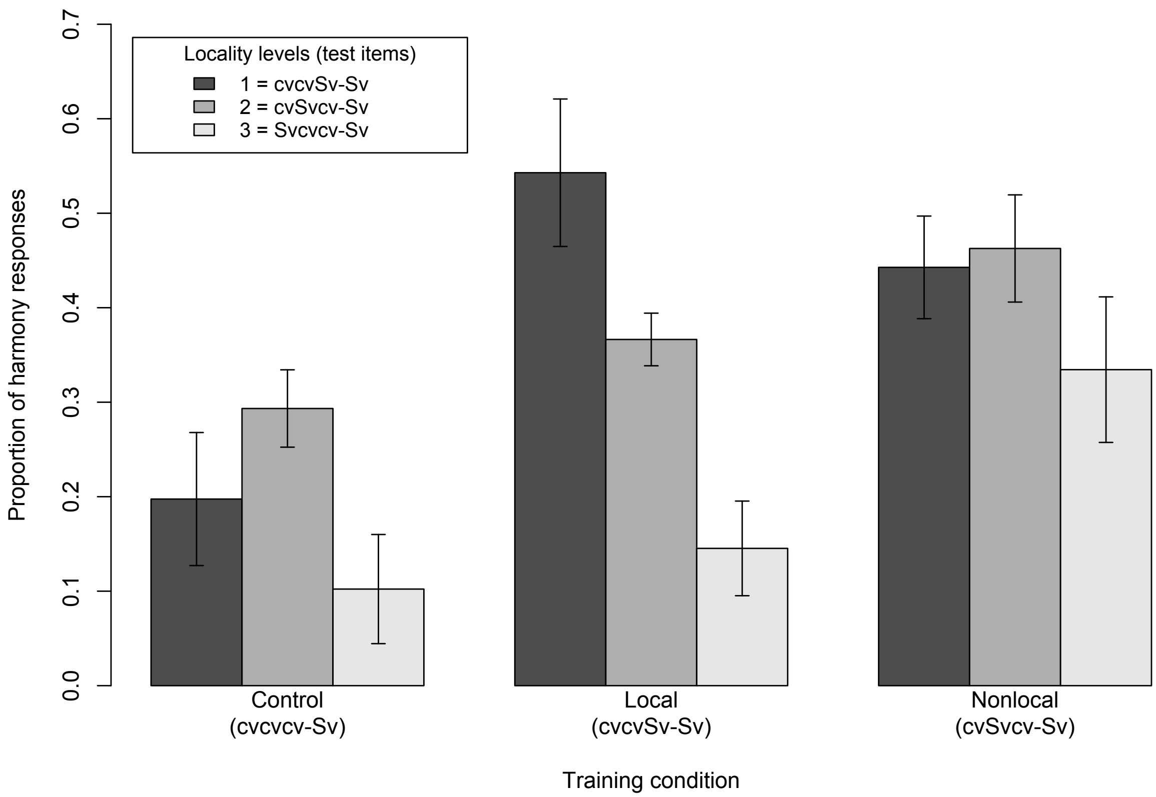


Figure 1 – Proportion of harmony responses for each training condition at each distance. Error bars represent the standard error, based on the means of subjects in that group.

The above results can be summarized as follows. Subjects in the control group are most likely to choose an alternation that results in harmony for sibilants in the second syllable of the root, and are least likely to choose an alternation for word-initial sibilants. Compared to the control group, subjects who were trained locally are more likely to choose an alternation in all positions, though this is most evident in the local cases, followed by the cases with a sibilant in the second syllable, with minimal distinction in the word initial position (at distance 3). Subjects in the nonlocal training condition are also more likely to choose an alternation at all distances compared to the control group. Compared to the local group, they are less likely to do so at distance 1, but more likely at distances 2 and 3.

The above results are interesting, but it is not yet clear which of these differences is meaningful. For example, if we determine that those trained in the local condition have indeed learned the local pattern, can we conclude whether or not they have generalized to the nonlocal items at distance 2? Similarly, those in the nonlocal condition appear to have learned the pattern at the distance they were trained on (distance 2), and generalized to the local contexts, but do we have evidence that they have generalized to the word initial positions at distance 3?

3.2 A mixed-effects logistic regression model

An alternative way to analyze data sets comprised of categorical responses, rather than using the average proportions for each subject, is with a logistic regression model (see Jaeger 2008 for a description of these models intended for linguists). In a logistic regression, the model finds the best fit for the log odds of choosing one response or another – in this case harmony vs. no harmony – based on a set of predictor variables. The predictor variables here, which are also categorical values, are the training conditions (Control, Local, Nonlocal), as well as the distance between sibilants for each testing item (1, 2, or 3). Note that the hypothesis predicts different results for each group at each distance, so it is important to include interactions between group and distance in the model as well. Other variables (including trial number, whether the test item with harmony was presented first or second, and whether the triggering suffix was -ʃi or -su) were included in other models, but showed no significant effects and the models including these variables did not give a significantly better fit to the data, so they are not included below in Table 1. Additionally, with the use of a mixed-effects logistic regression, the model can account for tendencies for individual subjects[2]. Barr *et al.* (in press) argues for a maximally rich random-effects structure in regression, so long as it is justified by the design. In the model presented below, a random intercept as well as a random slope for each distance was included for each subject. The model below was created with the *glmer* function from the *lme4* package (Bates *et al.* 2012) implemented in R (R Core Team 2012).

Table 1
A summary of the fixed portion of mixed-effects logistic regression model
(*N* = 827, # of Subjects = 30, log-likelihood = –456.8)

Coefficient	Estimate	SE	Pr(>|z|)
Intercept	-1.623	0.363	<0.001*
Distance 2	0.703	0.393	0.074
Distance 3	-1.171	0.536	0.029*
Local Group	1.802	0.481	<0.001*
Nonlocal Group	1.359	0.483	0.005*
Distance 2 x Local Group	-1.444	0.517	0.005*
Distance 3 x Local Group	-1.110	0.692	0.109
Distance 2 x Nonlocal Group	-0.6336	0.519	0.222
Distance 3 x Nonlocal Group	0.5212	0.658	0.428

In the table above, the estimate for the Intercept indicates the model's guess for the likelihood (in terms of log odds) that an average subject in the control group will choose harmony when presented with a distance 1 testing item. The negative estimate of -1.623 indicates that the model predicts the odds of choosing harmony to be $e^{-1.623} = 0.197$, which, in terms of probability, translates to a 16.5% chance. The estimates for the subsequent main effects of predictor variables indicate whether there is an increase or decrease in the likelihood of choosing harmony, as compared to this baseline.

With respect to the main effects, the estimate for distance 2 is positive, indicating an increase in the likelihood of control subjects choosing harmony, while the estimate for distance 3 is negative, indicating a decrease. Note that the effect of distance 3 reaches a significance level of <0.05. For distance 2, it is approaching significance with $p=0.074$.[3] For the main effects of group, both the Local and Nonlocal groups have positive estimates that are significant, and so they are much more likely to choose harmony at distance 1. This indicates that they have learned a consonant harmony pattern, or at least have begun to learn it. For the local group, this is the pattern they were trained on, but for the Nonlocal group, this effect demonstrates that they are choosing harmony at distance 1, even though they were trained at distance 2.

For the interactions of distance and group, the estimates indicate whether the likelihood of choosing harmony increases or decreases after already taking into account the main effects of group and distance, as well as the intercept. For the Local group, there is a significant decrease in the likelihood of choosing harmony at distance 2 and a decrease that is approaching significance at distance 3. This

[2] The results for individual subjects are reported in Appendix 1. They are quite interesting, as it seems that some subjects learn the pattern quite well, while others perform at a lower level than the control group. The reasons for and consequences of having different types of human learners have not been discussed in the literature, but this topic warrants further research.

[3] One problem with logistic regression models is that it can be difficult to assess significance in models that include interactions of the predictor variables. For this reason, the reader should not be concerned that some effects do not quite reach the arbitrary significance level, and instead focus on the effect size and whether it is positive or negative.

indicates that they have not generalized the local pattern that they learned to either of the nonlocal distances. In contrast, the interactions of the Nonlocal group with both distances 2 and 3 did not result in a significant increase or decrease in the likelihood of choosing harmony indicating that the pattern they have learned applies to all distances.

These estimates can be difficult to interpret because of their complexity. For interactions, the estimates do not represent a final probability, but a number that needs to first be combined with several. However, it is possible to calculate the final probabilities for each group at each distance. As an example, the overall odds of an average subject in the local group selecting harmony at distance 2 can be calculated as $e^{-1.623+0.703+1.802-1.444} = 0.570$, which translates to a 36.3% chance. The fitted probabilities for each group at each distance are presented in Table 2 below. They correspond roughly to the average proportions of harmony responses that were presented above in Figure 1, but are actually the resulting probabilities for the best fit from the logistic regression model after building in the random effects.

Table 2
Fitted probabilities of choosing harmony, based on the mixed-effects logistic regression model from Table 1

	Distance		
Group	1	2	3
Control	16.4%	28.5%	5.8%
Local	54.4%	36.3%	10.9%
Nonlocal	43.4%	45.1%	28.6%

4 Discussion

4.1 Comparing results to typology

The results presented above provide evidence in support of the hypothesis that humans will learn new phonological patterns and generalize them in a way that matches up to the typology of the world's languages. As described in section 1, the typology of consonant harmony reveals just two types of languages with respect to locality. The first type consists of the languages that apply harmony only when the two dependent consonants are separated by at most one vowel. The second type is the set of languages that apply harmony whenever the relevant consonants co-occur within some domain, as defined without reference to any phonological representation, including transvocalic contexts as well as longer distances. In the results of this experiment, we see that humans learning a new artificial language can learn both types of languages. However, it is how the learners tend to generalize the pattern that makes this result most interesting.

Consider the learners in the local group. Their relevant training items were restricted to instances of harmony applying in transvocalic contexts. Not only were they more likely than the control group to choose a harmonic response in the same types of items that they were trained on, but they did not (or at least were less likely to) generalize this to nonlocal distances. Once the learners in the local group discover the pattern, they do not extend it to nonlocal distances, even though such a pattern is attested in natural languages, and they saw no evidence to the contrary. That is, the subjects seem to be making the most direct application of their training as is possible – harmony occurs in local contexts, and nowhere else.

We now turn to the nonlocal group, whose relevant training items were restricted to instances of nonlocal harmony applying over two syllables (cvSvcv-Sv). Subjects in this group were more likely than the control group to apply harmony in the same types of words that they were trained on. Furthermore, they generalized this pattern to types of words that were not encountered in their training data – to roots that had a sibilant in the final syllable (a shorter distance 1), and roots that had a sibilant in the first syllable (a longer distance 3). So in this case, subjects discovered a pattern that applied harmony at a nonlocal distance. Additionally, they correctly rule out the possibility of it being the first type of language that applies harmony only at local distances, and so they settle on generalizing it to all contexts within the word, in line with the pattern we see in the second, and only other type of attested consonant harmony language.

4.2 A learning bias

An increasingly common explanation for why some phonological patterns are more common than others is that the more frequent patterns are more likely to survive over time because of cognitive learning biases. This type of bias, which facilitates the learning of some patterns, prevents the learning of others, and helps determine how a pattern is generalized, is labeled as an *analytic bias*. Over time, a propensity of humans to learn a certain pattern would result in the pattern having a high survival rate (Moreton 2008). This is different from a more traditional approach to phonological change that is based on phonetic naturalness, or *channel bias* – the tendency for some sounds to be either misproduced or misperceived in certain contexts. The result discussed above suggests that humans have a learning bias that affects how humans learn the properties of locality for long-distance interactions. The subjects, who had no experience with a language containing harmony, were able to learn the pattern and they generalized in a way that matched the typology. This is taken as evidence that it is a learning bias that has resulted in only two types of consonant harmony patterns in the world's languages.

4.3 Formal models and other possible outcomes

The results of this experiment support the hypothesis that humans will generalize a phonological pattern in a way that mirrors the typology of that pattern. However, note that the subjects in this experiment spoke no languages other than English, and presumably know nothing about the typology of consonant harmony systems. This section then, will ignore the typological facts and consider the possible outcomes from a learner that has no preconceived notion of how consonant harmony patterns should apply.

First, consider a learner that is trained only on items that exhibit local harmony, and some viable strategies that the learner might use to discover a consonant harmony pattern. One possible type of learner is one that relies on brute force, looking for a dependency between any two consonants in a word, no matter how far apart they are and without keeping track of distances. This learner, though trained only on harmony in local contexts, will learn only that there is a dependency between [s] and [ʃ]. When presented with testing items, it will incorrectly apply this pattern to both local and nonlocal sibilant pairs, since it does not record the distance. In order to prohibit it from overextending the pattern, there are at least two possible modifications that can be made. First is to allow the learner to also keep track of the distance between the two consonants. In the training phase, the learner will discover that a dependency exists between sibilants that are separated by a vowel. In the testing phase, it will look for any such instances and apply harmony. A second, similar option is to have the learner restrict its maximum search space to the largest distance between sibilants seen in the training data. It will have completed the training phase having never seen any instances of two sibilants that are separated by more than one vowel. During the testing phase, it will not even look for sibilants in a nonlocal relationship, and so it will apply harmony in the local cases only. Either of the latter two learners would give a result analogous to what we observed in the experiment for the local group.

With this in mind, however, a problem arises when considering the task of learning the pattern with input restricted to harmony at distance 2. Only the first of the three learners presented above, which incorrectly generalized a local pattern to nonlocal contexts, is capable of learning and generalizing in the same way as the nonlocal group. In this case, the modifications that solved this problem for the case of local harmony would result in it learning to apply harmony either at distance 2 only, or to apply it at any distance up to 2. This paradox emphasizes the fact that even though humans learners tend to generalize consonant harmony patterns in a way that directly relates to the typology, finding one learning algorithm that can simultaneously learn both types of languages is not trivial.

4.4 The precedence model of learning long-distance dependencies

In an attempt to explain how humans might learn long-distance dependencies, including consonant harmony, Heinz (2010) proposes a precedence model of learning long-distance phonotactics. In principle, it is a very simple model, and in practice it makes strong predictions about what the typology of consonant harmony should look like if language learners use the proposed strategy when learning the pattern.

In the precedence model of learning, a learner takes into account *precedence relationships* in addition to keeping track of bigrams. For example, in a word like "pants" /pænts/, the learner keeps track of the bigrams {p,æ},{æ,n},{n,t}, and {t,s}, which are all adjacent phonemes. Additionally, the learner tracks the precedence relationships {p,æ}, {p,n},{p,t},{p,s},{æ,n},{æ,t},{æ,s},{n,t},{n,s} and {t,s} without reference to distance, just that the first sound precedes the second somewhere in the word. Therefore, such a learner makes a distinction between co-occurrence restrictions between adjacent segments, and dependencies that hold between any two segments. For a case of nonlocal sibilant harmony (that is purely phonotactic), the precedence relations {s,ʃ} and {ʃ,s} would never be encountered in the language, so the learner would recognize that two different sibilants should no co-occur within a word. Note that this model does not keep track of any segments that occur in between the segments of each precedence relationship, and thus no intervening segment can have any influence on the nature of the precedence relationship. Such a property makes the strong prediction that there is no possibility of an intervening segment blocking the interaction between non-adjacent phonemes, or at least that such a property would never be learned. Heinz argues that this property is desirable since there are no discovered instances of blocking in the typology of consonant harmony. However, several languages, including some Berber dialects (Elmedlaoui 1995; Hansson 2010b) and Rwanda (Walker and Mpiranya 2005), may exhibit the blocking of consonant harmony, so a reevaluation of this aspect of the precedence learning model is in order.

There is another prediction that the precedence learning model makes by adding the notion of a consonant tier. By doing so, the learner can also keep track of bigram relationships between consonants alone, therefore making a distinction between consonants that are adjacent on the consonant tier, and consonants that are in precedence relations, but ignoring any further distinction regarding distance. This would give rise to the typological split mentioned above between languages that have only local dependencies, and languages in which the dependency holds at all distances. One downside to this model is that it does not account for the effects of similarity. The learner is equally capable of picking out dependencies among sounds that are relatively dissimilar, which is reflected neither in the typology of consonant harmony, nor in the experimental research reported in the literature. However, this algorithm is not incompatible with other models of learning that are biased towards picking out similar segments (e.g. Hayes and Wilson 2008), and such a combination could provide a more comprehensive model of how humans are learning phonological patterns.

4.5 Directions for future research

While this study has presented evidence that humans learn and generalize patterns in a way that reflects the patterns that occur in natural languages, we do not know the limits on human learning for unattested patterns. For example, is it possible to learn a distinction between sibilant harmony that occurs between sibilants across one vowel or two vowels, but not at further distances? Is it possible to learn a non-local sibilant harmony pattern that does not apply at local distances as well? Answers to these types of questions will help us to better understand the restrictions on human learning and whether there is a direct relationship with the typological distribution of phonological patterns.

Another question that merits investigation is whether evidence of a local dependency in the input would actually help a learner to discover the dependency at a greater distance. That is, would the learner first discover the (more typologically common) local dependency and then check for the dependency at the longer distances? This line of research would give us insight into the strategies that learners use when they are trying to discover the phonological patterns of a language.

5 Summary and conclusions

The goal of this study was to give an explanation for why the typology of consonant harmony, with respect to locality, includes only two types of languages – those with local harmony, and those with harmony that applies across all distances. A logistic regression mixed model showed that learners trained on sibilant harmony only at a local (transvocalic) distance learned the pattern, but did not generalize the pattern to sibilants at nonlocal distances (two or three syllables away from the triggering suffix). Learners trained on sibilant harmony only at the nonlocal distance (two syllables away) learned this pattern, and generalized not only to the shorter, local distance, but also to the word initial sibilants that were three syllables away from the triggering suffix. This experiment, with several differences in methodology, has replicated findings from Finley (2011, 2012). The results are taken as evidence that humans have a learning bias that affects how we learn and generalize patterns, and that this learning bias is responsible for the typological shape of consonant harmony systems.

References

Barr, D.J., R. Levy, C. Scheepers, and H.J. Tily. (in press). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*.

Bates, Douglas, Martin Maechler, and Ben Bolker. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. http://CRAN.R-project.org/package=lme4.

Elmedlaoui, Mohamed. (1995). *Aspects des respesentations phonologiques dans certains languages chamito-semitiques.* Rabat: Faculté des Lettres et des Sciences Humaines. [Doctoral dissertation, Université Mohammed V, 1992.]

Finley, Sara. (2011). The privileged status of locality in consonant harmony. *Journal of Memory and Language* 65: 74–83.

Finley, Sara. (2012). Testing the limits of long-distance learning: Learning beyond the three-segment window. *Cognitive Science* 36: 740–756.

Finley, Sara, and William Badecker. (2009). Artificial language learning and feature-based generalization. *Journal of Memory and Language* 61: 423–437.

Hansson, Gunnar Ólafur. (2001). *Theoretical and typological issues in consonant harmony*. Doctoral dissertation, University of California, Berkeley.

Hansson, Gunnar Ólafur. (2010a). *Consonant harmony: long-distance interaction in phonology*. Berkeley, CA: University of California Press.

Hansson, Gunnar Ólafur. (2010b). Long-distance voicing assimilation in Berber: spreading and/or agreement? *Actes du Congrès de l'ACL 2010 / 2010 CLA Conference Proceedings*, ed. by M. Heijl. Association canadienne de linguistique / Canadian Linguistic Association. Available online at: [http://homes.chass.utoronto.ca/~cla-acl/actes2010]

Hayes, Bruce, and Colin Wilson. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39: 379–440.

Heinz, Jeffrey. (2010). Learning long-distance phonotactics. *Linguistic Inquiry* 41 (4): 623–661.

Jaeger, T. Florian. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language,* 59: 434–446.

Moreton, Elliott. (2008). Analytic bias and phonological typology. *Phonology* 25: 83–127.

Moreton, Elliott. (2012). Inter- and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language* 67: 165–183.

Moreton, Elliott, and Joe Pater. (2011). Learning artificial phonology: a review. To appear in *Language and Linguistics Compass*. Available online at http://www.unc.edu/~moreton/Papers MoretonPater.Draft.3.5.pdf

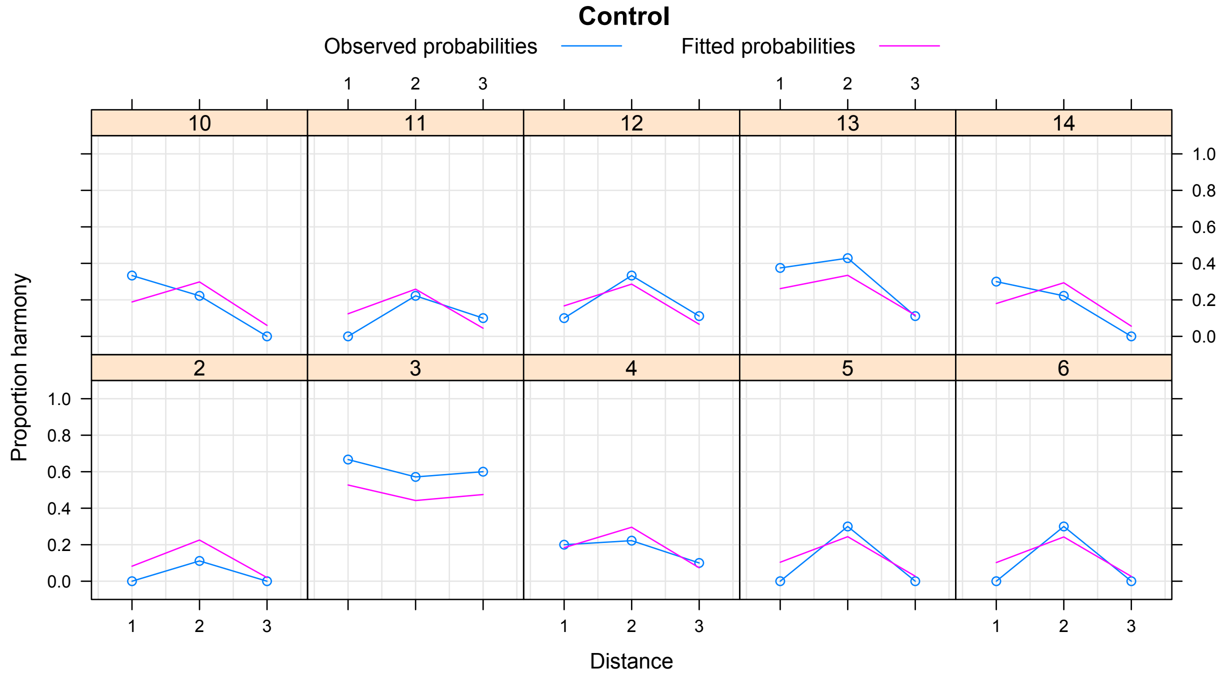
Nevins, Andrew. (2010). Two case studies in phonological universals: a view from artificial grammars. *Biolinguistics* 4: 218–233.

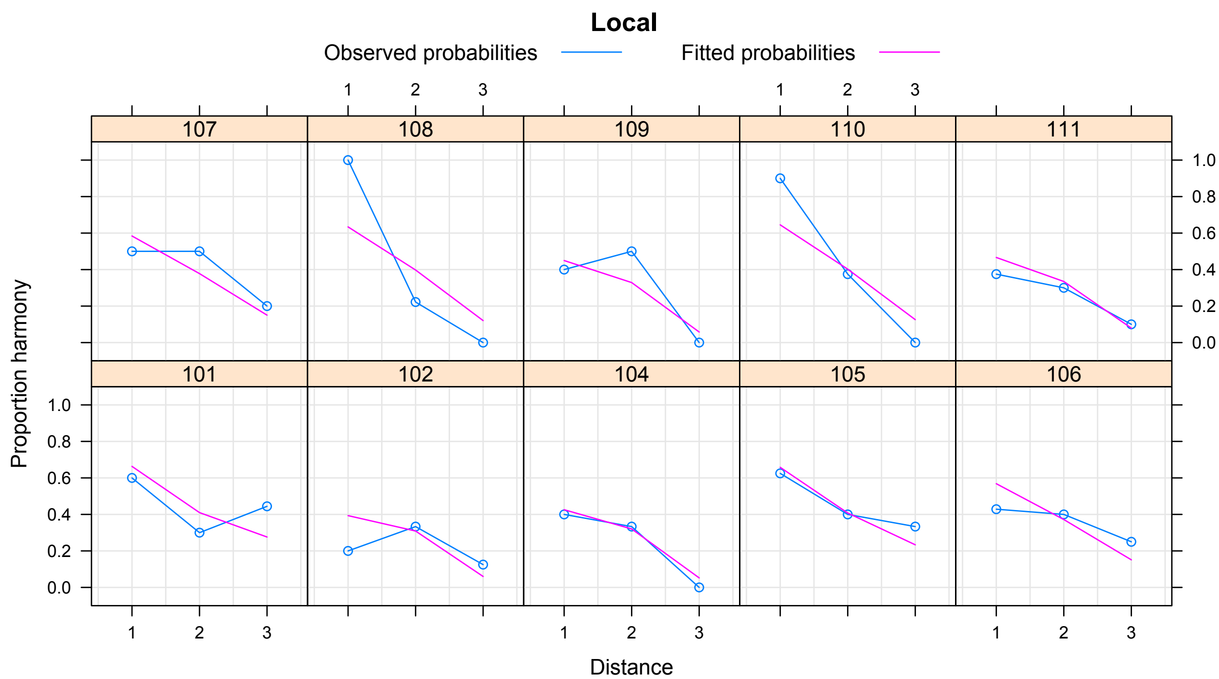
R Core Team. (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

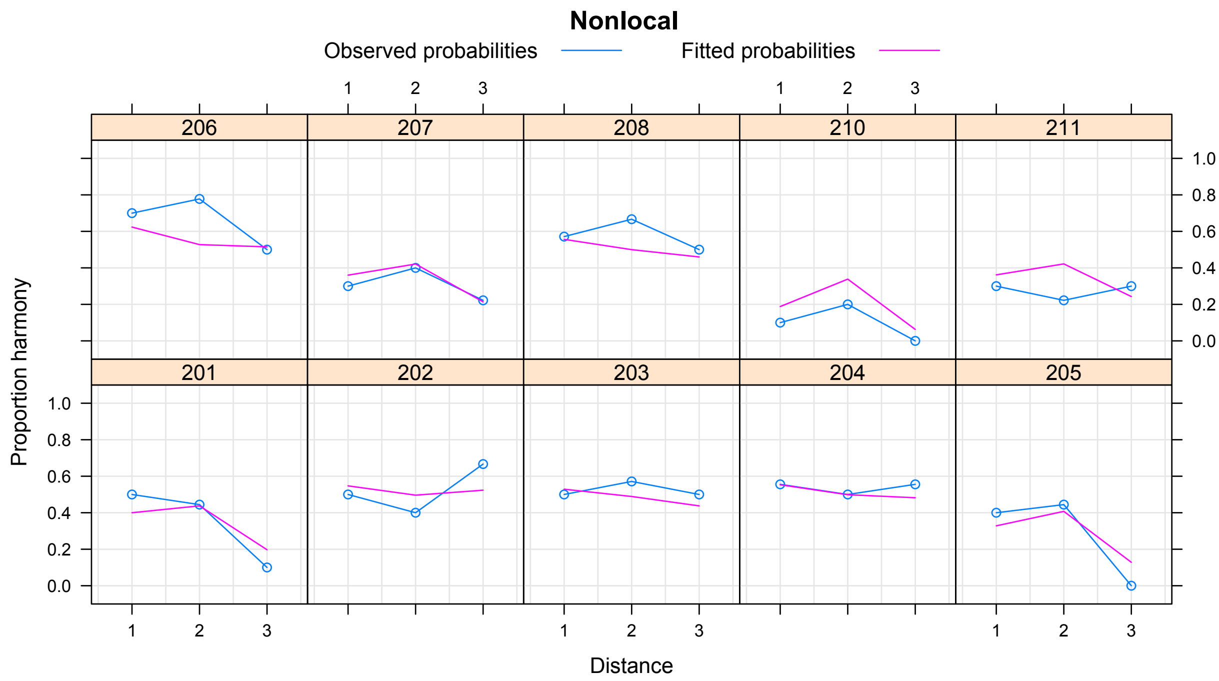
Rose, Sharon, and Lisa King. (2007). Speech error elicitation and cooccurrence restrictions in two Ethiopian Semitic languages. *Language and Speech* 50: 451–504.
Rose, Sharon, and Rachel Walker. (2004). A typology of consonant agreement as correspondence. *Language* 80: 475–531.
Walker, Rachel, and Fidèle Mpiranya. (2005). On triggers and opacity in coronal harmony. *Proceedings of BLS* 31: 383–394.

Kevin McMullin
kevinmcm@alumni.ubc.ca

Appendix 1 – Individual subject responses






Nonlocal
Observed probabilities
Fitted probabilities
206
207
208
210
211
201
202
203
204
205
Proportion harmony
Distance
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3

Context-of-use of augmented and unaugmented nouns in Nata*

Joash Johannes Gambarage

The University of British Columbia

Numerous studies of determiner (D) systems have used diagnostic tests such as: deictic, (in)definiteness, and (non)-specificity, to determine the contexts in which Ds and DPs are used. I use (in)definiteness and deictic tests to present a preliminary investigation of the context of use of augmented and unaugmented nouns in Nata – a Bantu language spoken in Northwestern Tanzania. I show that pre-prefixed (augmented) nouns surface in a variety of indefinite contexts and thus definiteness cannot be their inherent semantic feature. I argue that Nata PPFs are weak Ds which do not show any contrast of (in)definiteness and whose construal depends on the interaction between context-of-use and various operators or modifiers. I show further that overt PPFs can be used in both proximal and distal spatial/temporal contexts and thus they do not encode deictic force. As such, this paper contributes to our knowledge of Ds in Bantu.

1.0 Introduction

In some Bantu languages, the pre-prefix is claimed to mark definiteness (Bleek 1869; Mould 1974, Progovac 1993; Krifka & Zerbian 2008). In these analyses, the PPF is equated to the English definite article 'the' while the absence of PPF is equated to the indefinite English article 'a'. I show that augmented noun phrases (DPs) surface in a variety of indefinite contexts and thus definiteness cannot be their inherent semantic feature. I argue that Nata PPFs are weak determiners, which do not show a definiteness/indefiniteness contrast and whose construal depends on the interaction between context-of-use and various operators or modifiers. Specifically, unaugmented nouns are interpreted with respect to operators (i.e. polarity operators such as negation, a wh-morpheme in interrogatives, and a positive polarity item '-nde'); and augmented nouns are interpreted with respect to quantifiers, and inherently definite material (such as object markers and deictic demonstratives). When augmented nouns are not interpreted with respect to such contexts, nouns become consistently ambiguous between definite and indefinite reading. I argue that uniqueness and familiarity (for singular entities) or maximality (for mass and plural), which are the hallmark of definiteness (Heim 1991; Matthewson 1998; 1999; Kratzer 1998; Krifka 2003; Ionin 2001; Lyon 1999; Lyons 2011), are asserted when an object noun co-occurs with the OM or when a noun is used with deictic demonstratives which are inherently definite (see Visser 2008; Bresnan and Mchombo 1987 for OM argument in Bantu).

This paper divides into three sections. Section 1 gives a morphological and syntax-semantics overview of the PPF in Nata. Section 2 presents diagnostics for PPFs where I argue that Nata overt PPFs are not deictics as they are not sensitive to a proximal-distal temporal or spatial distinction. I will also argue that overt PPFs are not inherently definite markers but rather are definite when used with definite operators. In Section 3, I present contexts in which non-overt PPFs occur and show that their interpretation depends on scope order and/or the material surrounding them in a sentence. Section 4 concludes.

1.1 What is the pre-prefix?

As with many Bantu languages, Nata (E.45, Guthrie 1948) has a set of noun prefixes (N-prefix) that mark the class of the noun stem (N-stem). For example, the N-stem *-ɣɛsɔ* 'knife' occurs with the class 7 prefix *ki-*, (1a).

* I am very grateful to my committee: Rose-Marie Déchaine, Doug Pulleyblank and Michael Rochemont for reviewing

In addition, prefixed N-stems occur with an element commonly known as a pre-prefix or augment[1]. For example, the prefixed N-stem *kí-ɣɛsɔ* 'C7-knife' occurs with the pre-prefix (PPF) *e-*, (1b).

(1)	kí-ɣɛsɔ C7-knife[2]	b.	**e**-kí-ɣɛsɔ PPF-C7-knife

The shape of the PPF is determined by rules of vowel harmony (Johannes 2007). That is, the PPF vowel harmonizes with the vowel of the N-prefix (i.e., by copying the features of the N-class vowel)[3].

1.2 The Distribution of Nata Pre-prefixes

In Nata, PPFs come in two forms: PPFs that occur with the phonological form and PPFs that do not occur with the phonological form[4]. I dub PPFs that occur with the phonological form "overt PPFs" and PPFs that lack the phonological form "non-overt PPFs". Non-overt PPFs have a more restricted syntactic distribution than overt PPFs as they can only appear in the c-command domain of operators such as negation, question morphemes, and the positive polarity item '-nde' 'certain'. Nata non-overt PPFs are polarity sensitive and they consistently require a c-commanding licenser. On the other hand, nouns obligatorily appear with an overt PPF in the following contexts: with cardinals, (i.e. 'one'), with existential constructions, with universal quantifiers (∀-quantifier) like -ɔsɛ 'every' (for plural nouns) and -ɔsɛ 'all' (for singular nouns), with the generic quantifier in characterizing statements, with deictics, with object markers (OM), and for new discourse entities. These contexts are summarized in chart (2).

[1] In a language like Luganda pre-prefixes are simply a vowel. For that matter they are referred to as 'initial vowels (Ashton 1954; Hyman and Katamba 1993). Since in Nata the PPF in class 10 is a CV form, I will use the labels 'pre prefix' and 'augment' interchangeably as these labels are commonly used by Bantuists (Maho 1999) and are more appropriate to Nata.

[2] The following abbreviations are used in this paper: & = the sentence is ambiguous; FUT =future tense; * = ungrammatical; # = infelicitous; ∅ = non-overt; OM = object marker; NEG = negation, PF = prefix or noun class prefix; C = noun class, PASS = passive, APPL = applicative; SM = subject marker; INF = infinitival; LOC = locative; PFV = perfect; HABT = habitual; PAST = past; IPFV = imperfective; NMLZ=nominalizer.

[3] While PPFs are a vowel for nouns in other classes, the class 10/11 PPF stands out from all other PPFs in a number of respects. It marks plurality i.e., a-m-borí 'goat' ~ caa-m-borí 'goats'. It bears a high tone i.e., a-Ø-ŋwiína 'crocodile' ~ cá:-Ø-ŋwiina 'crocodiles'. It has a CV form. It is phonologically unpredictable. It marks N-V agreement. Since these attributes are of prefixes, this suggests that class 10 PPF behaves like a prefix, hence it has a double nature: as a PPF and as a prefix.

[4] The Phonological Form (PF) in Chomsky's (1995) sense, which I assume here, is the level of representation at which only information relevant to the phonetic realization of the utterance is present. In Chomsky's sense, PF is derived from surface structure (SS) and is considered the interface with the articulatory-perceptual system.

(2) The Distribution of augmented and unaugmented Nata nouns

Contexts	Unaugmented N	Augmented N
Polarity Contexts		
Negation	✓	✗
Wh-questions	✓	✗
Positive polarity item -nde	✓	✗
Quantifiers		
Cardinality	✗	✓
∃-construction –ɲihɔ 'there is'	✗	✓
∀-quantifier -ɔsɛ	✗	✓
Generic quantifier	✗	✓
Definite material		
Demonstratives	✗	✓
Object Marker	✗	✓
Indefinite discourse contexts		
A new discourse context	✗	✓

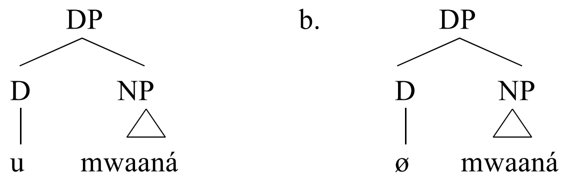
I argue that the presence of overt or non-overt PPF in (3) has a certain syntactic-semantic function, and thus PPFs are not semantically vacuous. I claim that PPFs share in common the denotation of an assertion of existence of an entity except where PPFs are used with operators whose semantics deny assertion of existence (i.e., negative polarity contexts). I analyze overt PPFs as weak Ds and non-overt PPF that feature in polarity contexts as "polarity weak Ds" (see also Matthewson 1999 on St'át'imcets, a Salish language in the same spirit). I argue that the interpretation of both overt PPFs and non-overt PPFs depends on the interaction with other local elements in a sentence and/or discourse contexts. Additionally, there is evidence that PPFs can interact in scope with logical operators such as negation, the question morpheme, and quantifiers (see Krifka and Zerbian 2008). In Nata, nouns used in these contexts are interpreted with respect to these operators. Furthermore, a DP with an overt PPF/D can be interpreted as definite if used with 'definiteness-possessing-material' such as deictic demonstratives or OM. Otherwise, the DP is neutral to definiteness if used with lexical units that have a neutral semantic feature to definiteness such as the cardinals (cf. Visser 2008 for similar argument on IsiXhosa). DPs which occur with the generic quantifier, i.e., the habitual morpheme in characterizing statements, or the positive polarity item **-nde** , 'certain' which possesses an indefiniteness value, will be construed as indefinite. I argue that the presence of an overt PPF on a DP does not guarantee the definite reading, since such DPs can easily surface in novel contexts, such as those introducing a new discourse referent.

1.3 The Syntactic Position of the Nata Pre-prefix

In order to meet descriptive adequacy and to correctly characterize and account for the syntax-semantic distribution of PPFs in Nata, I assume that in Nata, except for proper nouns, all noun arguments have a DP shell. PPFs sit in D^0 position and thus are D functional heads (cf. Visser 2008 for similar argument on IsiXhosa). The evidence for this comes from the fact that Nata quantifiers and demonstratives occur phrase-finally and thus they do not sit in D^0 position. I propose the DP internal structure for overt PPFs/Ds and non-overt PPFs/Ds for Nata as shown in (3)a and (3)b, respectively.

Fig 1: The Nata DP internal structure

(3) a. b.



In (3)a the D node is occupied by the overt PPF and in (3)b the D node is occupied by the non-overt PPF. Specifically, after spell-out the overt PPF in (3)a is syntactically located at PF and the non-overt PPF in (3)b is located at LF. The grammatical function of PPF/D (overt or non-overt) is assertion of existence except where a DP is used in negative polarity contexts whose semantics presuppose or assert existence. Since non-overt PPFs/Ds contribute to assertion of existence, I take this as evidence for a non-overt PPF over the absence of a PPF. For that matter, I abandon the usage of the term “unaugmented N” as it strikes me as confusing. Based on the distribution and the syntactic position of PPFs, I hypothesize the following about the Nata PPFs:

(4) i. PPFs are weak determiners which encode assertion of existence and they do not encode deictic force or definiteness.
ii. Overt PPFs do not encode definiteness unless they are used with deictics, OM or in contexts that are compatible with definite construal.
iii. In the absence of ‘definiteness-possessing-material’ (i.e. OM, and deictics), or indefinite modifiers (i.e., -*onde* ‘certain’), and context of use, DPs are ambiguous between definite and indefinite construal.
iv. Non-overt PPFs contribute to assertion of existence except where they are used in contexts whose semantics deny assertion of existence (i.e., negative polarity contexts).

2.0 Contexts of Use of Overt PPFs and their Interpretation

In this section, I discuss the contexts of use of overt PPFs and their syntax-semantics properties. I claim that overt PPFs are weak Ds. A weak determiner shows no deictic force, no definiteness/indefiniteness contrast, no specificity/non-specificity contrast (Matthewson 1998, 1999; Guillemin 2007, Lyon 2011). Since the determiner can be the locus of definiteness (Heim 1988), specificity (Enç 1991) or deictic force (Gillon 2006; Guillemin 2007) I employ definiteness and deictic force tests compiled from work such as Matthewson 1998, Krifka 2003, Heim 2011, Gillon 2006, Lyon 2011, Déchaine and Tremblay 2011, among others to argue for this position.

2.1 Testing for deictic force

In other languages the determiner is the locus for deictic force (see Gillon 2006 on Skwxú7mesh and Guillemin 2007 on Mauritius Creole). I present the data to show that Nata overt PPFs are not deictics, as they show no contrast between proximal and distal temporal/spatial demonstratives. Further, in the absence of demonstratives both the subject noun (5) and the object noun (6) are ambiguous between definite and indefinite reading.

(5) a. ***∅-mu-kári** a-kaa-n-dɔ́r-a
b. &***o-mu-kári*** a-kaa-n-dɔ́r-a
PPF-C1-woman SM3-PST-1sg-see-FV
(i) = a woman saw me’
(ii) = the woman saw me’

(6) a. *N-ka-rɔ́r-a ***∅-mu-kári***

b. &N-ka-rɔ́r-a *o-mu-kári*
1sg-PST-see-FV PPF-C1-woman
(i) = I saw a woman'
(ii) = I saw the woman'

Spatial deixis in Nata is fixed by the deictic demonstratives (7-8).

(7) a. ***∅-mu-kári** ú-nɔ a-kaa-n-dɔ́r-a
b. ***o-mu-kári*** ú-nɔ a-kaa-n-dɔ́r-a
PPF-C1-woman C1-this SM3-PST-1sg-see-FV
This woman saw me'

(8) a. *N-ka-rɔ́r-a ***∅-mu-kári*** ú-nɔ
b. N-ka-rɔ́r-a ***o-mu-kári*** ú-nɔ
1sg-PST-see-FV PPF-C1-woman C1-this
'I saw this woman'

In examples (7)-(8) both subject and object Ns can freely take proximal or distal spatial deictics. Spatial deictic demonstratives are generally considered to be inherently definite, (Lyons 1999; Krifka 2003; Visser 2008 etc). I argue that the presence of a deictic demonstrative induces the noun to be interpreted with a definite construal. Unlike deictic determiners in Skwxú7mesh (Gillon, 2006) and assertion-of-existence determiners in St'át'imcets (Matthewson, 1998), Nata PPFs do not encode deictic force or a visible/invisible contrast (see also Lyon 2011 for similar analysis on Okanagan). The overt PPFs in Nata can introduce a noun that is proximal to the speaker (9), distal (10) from the speaker, or totally invisible to the speaker (11).

(9) a. ***O-mu-kári*** a-ka-rɔ́r-a Luciá hánɔ
PPF-C1-woman SM1-PST-see-FV Lucy here
'The woman saw Lucia here'

b. Luciá a-ka-rɔ́r-a ***e-ɣí-taβo*** muu-mɛɛzá í-nɔ
Lucy SM1-PAST-see-FV PPF-C7-book LOC-C9table C9-this
'Lucy saw the book on this table here

(10) a. ***O-mu-kári*** a-ka-rɔ́r-a Luciá haarí / hááári
PPF-C1-woman SM1-PST-see-FV Lucy there/ way over there
'The woman saw Lucia there/ way over there'

b. Luciá a-ka-rɔ́r-a ***e-ɣí-taβo*** muu-mɛɛzá yiirí / yíííri
Lucy SM1-PST-see-FV PPF-C7-book LOC-C9table that / over there
'Lucy saw the book on that table/ that table way over there

(11) a. ***O-mo-súβe*** n-aa-ku-ɣɛ́ɛnd-a muu-mw-ɛɛrí
PPF-C1-person ?-SM3-PST-walk-FV LOC-C3-moon
A person is walking on the moon'

b. ***o-mu-chaánga*** w-o-ɲí muu-mw-ɛɛrí
PPF-C1-sand ?-SM3-there.is LOC-C3-moon
'There is sand on the moon'.

I show that temporal deixis is not encoded by Nata Overt PPFs. A noun with an overt PPF can refer an entity from the past (12), an entity in the present (13) or an entity in the future (14).

(12) a. *Kárɛ ∅-*βá-rem-i* m-ba-arɛ́ ko-rém-a o-βú-βaamba mo-o-ro-hóɣo
b. Kárɛ *a-βá-rem-i* m-ba-arɛ́ ko-rém-a o-βú-βaamba mo-o-ro-hóɣo
in.the.past PPF-C2-cultivate-NMLZ ?-C2-were C15-grow-FV PPF-c14-cotton LOC-PPF-C1-plain
'In the past farmers used to grow cotton in the flatland'.

(13) a. *čaa-∅-sikó chí-no ∅-*βá-rem-i* m-ba-haa-réma o-βú-βaamba mu-čaa-∅-kɛrɛ́ɣɛ
b. čaa-∅-sikó chí-no *a-βá-rem-i* m-ba-haa-réma o-βú-βaamba mu-čaa-∅-kɛrɛ́ɣɛ
PPF-C9-day C10-dem PPF-C2-cultivate-NMLZ ?-SM2-HABT-grow-FV PPF-C14-cotton LOC-PPF-C9-highland
'Nowadays farmers grow cotton in the highlands'.

(14) a. * čaa-∅-sikó čí-nɔ čee-kúu-č-a ∅-*βá-rem-i* m-ba-ko-rém-a
b. čaa-∅-sikó čí-nɔ če-kúu-č-a *a-βá-rem-i* m-ba-ko-rém-a
PPF-C9-day C10-dem C9-C15-come-FV PPF-C2-cultivate-NMLZ ?-SM-FUT-grow-FV

o-βu-βaamba mu-čaa-∅-taarâari
o-βu-βaamba mu-čaa-∅-taarâari
PPF-C14-cotton LOC-PPF-C10-valley
'In the coming days, farmers will grow cotton in the valleys'.

Since Nata deixis is fixed by demonstratives, and since there are no deictic restrictions on the use of an overt PPF on a noun, I take this as evidence that Overt PPFs in Nata are 'neutral to deixis' and that they are not specified for deictic force. I conclude that Nata Overt PPFs do not encode deictic force.

2.2 Testing for definiteness

The determiner can be the locus for definiteness (Heim 1988; Krifka 2003). In this section I show that Nata PPFs cannot be analyzed as definite Ds. PPFs do not assert uniqueness/familiarity and thus cannot be analyzed under a Russellian account where uniqueness is asserted (Russell, 1905). Also, PPFs do not presuppose uniqueness and thus cannot be couched in Fregean account where uniqueness/maximality is presupposed (Frege, 1892). It is generally acceptable that a presupposition of uniqueness and an assertion of uniqueness are hallmarks of definiteness. I discuss and then reject the semantic properties of definiteness (familiarity/uniqueness/ maximality), and specificity as either presuppositional or assertive components of the denotation of Nata PPFs.

2.2.1 Defining definiteness

In a language like English, the definite D *the* is associated with an iota operator (ι), (15) or with the Maximality operator (MAX), (16), for plural referents which when mapped to a predicate yield the greatest individual(s) (cf. Sharvy 1980; Krifka 2003:194; Heim 2011)[5].

(15) $[[\text{the}]] = \lambda P\text{: } \exists x \forall y\ [P(y) \leftrightarrow x = y].\iota x.P(x)]]$ (Heim 2011:194)

(16) $[[\text{the}]] = \lambda P\text{: } \exists x \forall y[MAX(P)(y) \leftrightarrow x = y].\iota x.\ MAX(P)(x)$

5 Mass nouns are not associated with the concept of specificity or uniqueness and instead are associated with concepts such as inclusiveness or totality or partiality (see Lyons 1999). In formal accounts definite mass nouns are interpreted within the restrictor of the Maximality operator (see Heim 2011).

The denotation in (15) refers to the familiar/unique referent in a possible world w, and (16) refers to the familiar/unique number of referents in a possible world w. On the other hand, the English indefinite article *a* is analyzed as an existential (∃) quantifier. For instance, the DP *a dog* in a sentence like *a dog ran* can be represented by (17).

(17) [[a dog]] = λP∃x[DOG(x) & P(x)]

The semantics in (17) talks about a non-familiar dog as opposed to a familiar/unique dog in (15). In English, the presence of overt determiners (*a, the*) restricts the free application of operators ∃ and ι, or (MAX) for plural referents, (see Krifka 2003). As one can see the distinction between definiteness (26-27) and indefiniteness (17) generally refers to familiar-novel distinction (see Heim 1991, Mathewson 1998, Lyons 1999). I take it that familiarity (and uniqueness) are definite features while novelty, non-familiarity, (and non-uniqueness) are indefiniteness features of Nata DPs. Here, definiteness requires both the speaker and the hearer to be familiar with the discourse referent.

2.2.2 PPFs do not presuppose uniqueness

In the following data I show that there is no familiarity requirement associated with the use of overt PPF. The use of augmented NPs does not impose any constraint on the common ground of either speaker or hearer. This is to say PPFs do not presuppose the existence of a unique referent, which satisfies the predicate. I give various examples that support this generalization. First, an augmented N can introduce a new discourse referent.

(18) a. *Hayo kárɛ n-aarɛ́-hɔ ***∅-mu-tɛ́mi***
b. Hayo kárɛ n-aarɛ́-hɔ ***o-mu-tɛ́mi***
there long.ago ?-be-LOC PPF-C1-chief
'Long ago there was a chief'/'Once upon a time there was a chief'

The use of overt PPFs in (18) patterns in the same way as the English indefinite *a*. However, the overt PPFs are also licit in discourse anaphoric contexts. In this case an augmented N is compatible with the definite reading in (20). Assume (20) is a continued story from (18) repeated in (19).

(19) a. *Hayo kárɛ n-aarɛ́-hɔ ***∅-mu-tɛ́mi***
b. Hayo kárɛ n-aarɛ́-hɔ ***o-mu-tɛ́mi***
there long.ago ?-be-LOC PPF-C1-chief
'Long ago there was a chief'/'Once upon a time there was a chief'

(20) a. ****∅-mu-tɛ́mi*** n-aarɛ́ mu-sáambaarok-u
b. ***o-mu-tɛ́mi*** n-aa-rɛ́ mu-sáambaarok-u
PPF-C1-chief ?-be-PST C1-charm-NMLZ
'The chief was charming'

In (20) the discourse context provides the value for the interpretation of an augmented N. Examples (19) and (20) show that it is felicitous to use overt PPFs in both definite and indefinite contexts. It is not surprising to encounter data from non-narrative contexts also showing that overt PPFs may be used in non-familiar and out-of-the-blue contexts, (21).

[Context: We meet on the way and I tell you what happened to me today].

(21) a. *Wéeche? rɛɛrɔ́ ***∅-moo-sirikaré*** a-kaa-ɲí-imeerer-i

b. Wéeche? rɛɛrɔ́ ***o-moo-sirikaré*** a-kaa-ɲí-imeerer-i
guess what? Today PPF-C1-police.officer SM-past-OM-stand-FV
'Guess what? Today, a police officer stopped me.'

In contexts such as (21) the hearer is not expected to be familiar with the particular police officer reported by the speaker. If PPFs were Fregean definites, we might expect presupposition failure in contexts where the police officer referred to here was not in the hearer's common ground, yet there is no presupposition failure for (21). In question-and answer contexts, an overt PPF may be used in answers that establish a new discourse referent (22), or answer a question related to a previously established discourse referent (23).

[Context: The questioner sees the addressee opening the cupboard with a variety of utensils, glasses, and cups and has no idea what the addressee wants to take from the cupboard].

(22) Q. What do you want?
a. *Ni-kwɛnd-a ***∅-ɣi-kɔ́ɔmbɛ***
b. Ni-kwɛnd-a ***e-ɣi-kɔ́ɔmbɛ***
1s-want-FV PPF-C7-cup
'I want a cup'.

[Context: The speaker is asking about a specific dog which is discourse old].

(23) Q. Where is the dog?
a. ****∅-∅-seesé*** ye-eɲí mw-∅-í-βendo
b. ***a-∅-seesé*** ye-eɲí mw-∅-í-βendo
PPF-C9-dog C9-be.at LOC-PPF-C5-backyard
'The dog is in the backyard'

If overt PPFs presupposed uniqueness, it would not be possible to have an augmented N felicitously used in an indefinite context such as in (22). This is further evidence that overt PPFs cannot be analyzed as definite. The tendency of overt PPFs to surface both in definite and indefinite contexts (cf. ex. (22) and (23) respectively) suggests that overt PPFs are weak Ds which depend on pragmatic and syntax-semantic factors for their interpretation.

I finally show that presupposition of uniqueness/maximality is possible when augmented Ns are used with operators that encode definiteness. The OM morpheme in Bantu is equated to the definite pronoun (Bresnan & Mchombo 1987). An object NP used with the OM denotes familiarity/uniqueness of the referent. Indices indicate coreference:

(24) a. *Witáre n-a-ko-**ɣí**$_i$-sɔm-a ***∅-ɣí-taβo***$_{i/*j}$
b. Witáre n-a-ko-**ɣí**$_i$-sɔm-a ***e-ɣí-taβo***$_{i/*j}$
Witare ?-3sg-IPFV-OM-FV PPF-C7-book
Witare is reading (it) the book.

The OM in (24) is in coreference with the object NP. In this case the book being talked about here is in the common ground of both the speaker and the hearer. It is possible to propose that the OM here exactly stands as an overt realization of an iota operator as proposed in works such as Chierchia (1998) and Krifka (2003). This follows straightforwardly from the fact that in most Bantu languages the OM possesses an inherent definite feature (see also Visser 2008 for a similar argument). A similar phenomenon is found with the Nata possessive morphology, which is plausibly tied to a presupposition of existence. In Nata a possessed N must be augmented. A possessed count N in Nata is definite, as in (25).

(25) a. ***∅-ɣi-kɔ́ɔmbɛ** ke Ɉaŋgí n-ke-eɲí mu-u-∅-mɛɛzá

b. ***e-ɣi-kɔ́ɔmbɛ*** ke Ɉaŋgí n-ke-eɲí mu-u-∅-mɛɛzá
PPF-C7-cup C7.POSS Nyangi ?-C7-be.at LOC-PPF-C9-table
'Nyangi's cup is on the table'

I conclude that Nata PPFs do not presuppose the existence of a unique referent, which would be predicted under the Fregean analysis of English as stated in (15) above. Concretely, Nata overt PPFs/Ds are not inherently definite. Rather, definiteness comes from sources such as OM, possessor morphology, the demonstratives, and the discourse context.

2.2.3 PPFs do not assert uniqueness

It is also plausible to test for definiteness using a Russellian account of assertion of uniqueness/maximality. I follow the spirit of various semanticists (Sharvy 1980, Heim 1988, 2011, Krifka 2003, Lyon 2011, and many others) in considering uniqueness assertions, and maximality assertions for plural DPs as hallmark properties of definiteness. If Nata PPFs were construed similarly to the English definite determiner *the*, Nata PPFs would crucially encode assertion of uniqueness, $\iota x.P(x)$. For some situations, this seems to be true since an augmented N may introduce singular definite referents, such as the Sun and Moon. The syntactic position plays no role here.

(26) a. ***∅-mw-ɛɛrí** o-ri-iβís-ire mu-u-ma-sáaro

b. ***u-mw-ɛɛrí*** o-ri-iβís-ire mu-u-ma-sáaro
PPF-C5-moon SM3-C5-hide-PFV LOC-PPF-C6-cloud
'The sun's hidden behind the clouds'

(27) a. *a-βa-náata m-ba-haa-sáásaam-a ***∅-ry-ooβá***

b. a-βa-naata m-ba-haa-sáásaam-a ***i-ry-ooβá***
PPF-C2-Nata ?-C2-HABT-worship-FV PPF-C5-sun
'Nata people worship the sun'

Evidence for the fact that PPFs do not assert uniqueness/maximality come from considering further Nata data presented below. First, I show that it is felicitous to use an augmented N in situations where there is more than one contextually-salient element satisfying the nominal property.

[Context: Several cups on a table, equidistant from speaker].

(28) a. *Nu-u-h-ɛ́ ***∅-ɣi-kɔ́ɔmbɛ***

b. Nu-u-h-ɛ́ ***e-ɣi-kɔ́ɔmbɛ***
1sg-2sg-give-FV PPF-C7-cup
'Give me a cup'

(29) a. *Nu-u-h-ɛ́ ***∅-βi-kɔ́ɔmbɛ***

b. Nu-u-h-ɛ́ ***e-βi-kɔ́ɔmbɛ***
1SG-2SG-give-FV PPF-C8-cup
'Give me some cups'

The response to a request made in (28) would be that the addressee would pass one of the cups to the speaker. This context is therefore incompatible with assertion or presupposition of uniqueness (cf. Lyon 2011 on Okanagan DP). Plural augmented Ns in Nata are also not definite. In example (29) the addressee would respond by either grabbing all the cups, or any number of cups or seek clarification from the speaker on the number of cups the speaker wants. This argues that assertion of maximality by an augmented N is not encoded. As with Nata augmented plural Ns, there is no restriction for mass Ns to surface with a PPF. Further evidence reveal that augmented mass nouns in Nata do not assert maximality of a referent. The context in (30) shows that the augmented mass N is not maximal.

[Your friend is preparing a relish. There was a container of salt on the table, but now it's gone. I ask "What happened to the salt?" Your friend replies:]

(30) a. ****∅-moo-ɲó*** n-né-toor-ire mɔ-ɔ-kɛ-rɛ́ɛrɔ
b. ***o-moo-ɲó*** n-ne-toor-ire mɔ-ɔ-kɛ-rɛ́ɛrɔ
PPF-C3-salt ?-1sg-C3-put-PFV LOC-PPF-relish

o-woo-ndé n-né-**wo**-tor-ire mu-u-∅-kabati
o-woo-ndé n-né-**wo**-tor-ire mu-u-∅-kabati
PPF-C9-another ?-1sg-C3-put-PFV LOC-PF-C9-cupboard
'I put (some) salt in the relish, the rest I put (it) in the cupboard'.

c. #***o-moo-ɲó*** n-né-**wo**-toor-ire mɔ-ɔ-kɛ-rɛ́ɛrɔ
PPF-C3-salt ?-1SG-C3-put-PFV LOC-PPF-relish

o-woo-ndé n-né-**wo**-tor-ire mu-u-∅-kabati
PPF-C9-another ?-1sg-C3-put-PFV LOC-PF-C9-cupboard
'#I put (it) the salt in the relish the rest I put it in the cupboard.'

As with all augmented Ns, the augmented N in (30)b asserts the existence of a referent but it does not assert maximality. The object N here is interpreted in relation to the weak quantifier 'some' even though the object N is not quantified. This follows straightforwardly from the analysis of Gambarage (2012) who argues that augmented Ns in Nata are in most cases interpreted within the restrictor of the existential quantifier if they are not within the restrictor of any other logical operator in a sentence. Further, (30)b cancels any maximality implicature that the salt/ all the salt was put in the relish. If the co-occurrence of the object N with OM asserted maximality we would predict that this sentence would be infelicitous just like in the English gloss *#I put the salt in the relish and the rest I put it in the cupboard',* and correctly this is the result we see in (30)c**.** Likewise, plural augmented count Ns do not assert maximality. In (31) maximality implicature of the sentence is easily cancellable.

[Context: The students are writing their music theory exam from 10.00am-12.00; it's now 11.00am.]

(31) a. ****∅-βaa-nafuunzí*** βa-ɣa-sám-a ɣu-kɔ́r-a ii-ɣɛ́ɛmbɛ e-me-tiháani saa-nnɛ́
b. ***a-βaa-nafuunzí*** βa-ɣa-sám-a ɣu-kɔ́r-a e-me-tiháani saa-nnɛ́
PPF-C2-student SM2-PST-start-FV INF-do-FV PPF-C3-exam time-4(=10)

a-∅-taβɔ́ɔri a-βa-mwé βa-ɣa-téera
a-∅-taβɔ́ɔri a-βa-mwé βa-ɣa-téera
PPF-C9-morning PPF-C2-some SM2-past-be.late
'(Some) students started to write the exam at 10.00am in the morning (but) some were late'.

If an augmented N asserted maximality in this case, the sentence in (31) would implicate that all the students started to write the exam at 10.00am. It is plausible to conclude here that overt PPFs assert existence but do not assert maximality.

The last piece of evidence I consider here comes from characterizing statements. Characterizing (generic) sentences express generalizations about sets of entities or situations (Krifka 2003; Déchaine and Tremblay (2011). Such statements are interpreted within the restrictor of the covert generic operator (GEN), proposed by Chierchia (1998) and Krifka (2003), which usually select for the generic reading. In Nata generic statements require the habitual aspectual marker, which I argue is the overt realization of GEN (see also D&T 2011 on Shona)[6]. In (32) the NP does not refer to a specific referent but still it appears with an overt PPF.

(32) a. ***∅-*mo-rísi*** n-a-haa-séeɣa cáá-∅-ŋɔmbɛ
b. ***o-mo-rísi*** n-a-haa-séeɣa cáá-∅-ŋɔmbɛ
PPF-C1-shepherd ?-SM1-HABT-like-FV PPF-C9-cattle
'A shepherd likes cattle'/ *The shepherd likes cattle'.

If the use of overt PPFs encoded definiteness in all contexts, then we would not expect example (32) to be felicitous. If we mark (32) with a different tense/aspect it will yield a different construal. This suggests that there is an interaction between temporal orientation and the interpretation of overt PPFs. I conclude this section by asserting that overt PPFs in Nata are consistent with assertion of existence.

3.0 Context of use of Non-overt Pre-prefixes

As with overt PPFs, I argue that non-overt PPFs contribute to assertion of existence and are also interpreted with respect to material surrounding them in a sentence. Specifically, non-overt PPFs assert no existence when only used with local elements whose semantics are incompatible with assertion of existence (i.e., in negative polarity contexts) but assert existence when used with materials that are consistent with assertion of existence (i.e., the positive polarity item '-nde' 'certain'-the indefiniteness modifier). DPs with non-overt PPFs/Ds may take wide scope with respect to negation where being in such position they consistently assert existence. In spite of the fact that "wide-scope-non-overt PPFs" behave more like wide-scope indefinites in scopal behaviors, Nata non-overt PPFs cannot be fully analyzed as English-style indefinite or non-polarity wide scope St'at'imcets Ds. One major difference between Nata wide-scope-non-overt PPFs and English-style indefinite or non-polarity St'at'imcets indefinites is that wide-scope-non-overt PPFs rely heavily on pragmatic factors for their interpretation. I now turn to discuss non-overt PPF cases.

Non-overt PPFs are scopally ambiguous when used with certain logical operators. Example (33) indicates that when a DP occurs in contexts such as negation, it is c-commanded by the negation and hence it cannot occur with an overt PPF. Depending on the scope order, the DP is compatible with the non-assertion of existence (polarity reading) and assertion of existence (specific reading).

(33) a. Musá ta-a-ɣor-íre **ø-ɣí-taβo**
b. *Musá ta-a-ɣor-íre **e-ɣí-taβo**
Moses NEG-PST-buy-PFV PPF-C7-book
(i) 'Moses didn't buy a book'
(ii) 'Moses didn't buy any book'

I argue that in (33) when the DP takes lower scope with respect to the NEG it yields the non-assertion of existence (polarity) reading, which is compatible with the context given for example (34) below.

[Context: Students visited a bookshop. It was their choice either to buy or not to buy books. Some students bought some books but Moses didn't. One student says (34)].

(34) a. Musá ta-a-ɣor-íre ***∅-ɣí-taβo***
b. *Musá ta-a-ɣor-íre ***e-ɣí-taβo***

[6] I thank Rose-Marie Déchaine (p.c) for bring this issue to my attention.

Moses NEG-PST-buy-PFV PPF-C7-book
'Moses didn't buy any book'
i. ✔Moses didn't buy any book
✔¬ [∃x [book (x) & [buy (x) (Musa)]]
ii. ≠ There is a book that Moses didn't buy (unavailable reading)
≠∃x [book (x) & ¬ [buy (x) (Musa)]

In (34) the D asserts no existence of a book that Moses bought. I analyze a non-overt PPF in (34) as a negative polarity D and not as negative polarity item (NPI) as Progovac (1993) proposes for Kinande. There are two reasons for not extending the Kinande analysis into Nata. One reason is that Nata has an overt NPI *'ɔ-ɔsɛ'* which can be felicitously used in contexts such as (34) repeated in (35).

(35) a. Musá ta-a-ɣor-íre **∅−ɣí-taβo ky-ɔɔ-ky-ɔ́ɔsɛ**
b. *Musá ta-a-ɣor-íre ***e-ɣí-taβo ky-ɔɔ-ky-ɔ́ɔsɛ***
Moses NEG-PST-buy-PFV PPF-C7-book C7-RED-C7-any
'Moses didn't buy any book'
i. ✔Moses didn't buy any book
✔¬ [∃x [book (x) & [buy (x) (Musa)]]
ii. ≠ There is a book that Moses didn't buy (unavailable reading)
≠∃x [book (x) & ¬ [buy (x) (Musa)]

Both (34) and (35) do not allow the wide scope reading with respect to the negation hence non-assertion of existence (polarity) reading. The second reason is that the non-use of the overt NPI allows the emergence of a wide scope reading (ex. (33) also ex. (36)) while the use of the overt NPI would always block such a reading (ex. (35) above).

In 36(c) I show that the overt NPI is illicit when it is used in the assertion of existence context as it blocks such reading just like in example (35).

[Context: Students visited a bookshop. Every student was supposed to buy an English textbook by the author Mangi. Moses didn't have any money. Students are in class and one student says:] . (Adopted from Matthewson (1998: 95-97)).

(36) a. Musá ta-a-ɣor-íre ***∅-ɣí-taβo***
b. *Musá ta-a-ɣor-íre ***e-ɣí-taβo***
Moses NEG-PST-buy-PFV PPF-C7-book
'Moses didn't buy a book'
i. ✔There is a book that Moses bought
✔∃x [book (x) & ¬ [buy (x) (Musa)]
ii. ≠ Moses didn't buy any book
≠ ¬ [∃x [book (x) & [buy (x) (Musa)]]

c. ≠Musá taa-ɣor-íre ***∅−ɣí-taβo ky-ɔɔ-ky-ɔ́ɔsɛ***
≠ 'Moses didn't buy any book'

In (36) the DP is interpreted with a wide scope over the NEG operator to create a specific reading. Here, Nata presents a different system than that described in Kinande by Progovac (1993). The point of departure between the overt NPI and the non-overt PPF/D is that the overt NPI blocks the wide scope reading (quantificational interpretation) of the DP while the non-overt PPF/D allows the wide-scope interpretation. The non-polarity (wide scope/assertion of existence) D here parallels the English indefinite D in a sentence like *Moses didn't buy a book* in taking a wide scope reading[7]. Concretely, I argue that in (36)a the non-overt PPF is syntactically located in the Logical Form.

The Nata positive polarity -*ndé* 'certain' is a specificity marker, which always forces a specific indefinite reading. When the -ndé modifier is used with a noun only the non-overt PPF is licensed. The DP used with this modifier encodes non-emptiness of a set.

(37) a. ***∅-mo-súβe wóo-nde*** a-hét-ire há-nɔ
b. ****o−mo-súβe wóo-nde*** a-hét-ire há-nɔ
PPF-C1-man SM1-certain SM3-pass-PFV C2-here
'a certain man passed here

I propose that the -ndé is an operator, which applies to the predicate to yield a specific individual corresponding to the speaker's description. It is evident that the non-overt PPF in example (37) corresponds to assertion of existence just like the wide scope assertion of existence Ds in negative sentences.

In Nata wh-questions, the question operator -ké licenses the non-overt PPF on a noun. However, the interpretation of a noun will depend on context of use. For instance, in nonfactual contexts the non-overt PPF yields a non-assertion of existence while in factual contexts it yields an assertion of existence.

38) a. ***∅−mo-súβe-ké*** u-mw-aaná a-tέm-irɛ?
b. ****o-mo-súβe-ké*** u-mw-aaná a-tέm-irɛ?
PPF-C1-man-WH PPF-C1-child SM1-beat-PFV
'Which man did a/the child beat?'

The wh-morpheme **–ké** in Bantu is claimed to posses an inherently lexical-semantic feature of indefiniteness (see Visser 2008). This is compatible with my analysis that the indefiniteness feature is not a property of non- overt PPFs. Rather the indefiniteness feature here comes from the wh-morpheme. Thus (38), if used in factual contexts, yields the specific reading compatible to the English sentence *"there is some man a/the child beat but I don't know which one it is".* The latter is compatible with Matthewson (1998 and references therein) who adduces that **wh**-questions presuppose the content of a non-questioned portion.

When a noun is used in polarity contexts in yes-no questions it can only surface with a non-overt PPF. For that matter in (39) the DP asserts existence. Otherwise, the overt PPF is illicit in non-polarity contexts such as (40).

7 Fodor and Sag (1982) dismiss the idea that indefinites in English are quantificational due to their tendency of being unconstrained by the island principle. I have not worked on sentences containing islands to see whether or not this is true of Nata. However, there is enough evidence that Nata overt PPFs are quantificational, i.e., in (1) when a quantified augmented N takes narrow scope it is compatible with the non-specific indefinite reading (i). On the other hand, when the lower augmented N scopes over the quantified DP it is construed as a specific indefinite, (ii).

(1) a−n-gúɣe yɔ-ɔsέ yee-ku-r-í **e-**ɣí-tɔkɛ
C9-baboon C9-every SM9-FUT-eat PPF-C7-banana
Every baboon is eating a banana
i. ✔Every baboon is eating its own banana
✔ $\forall y$ [baboon (y) $\rightarrow \exists x$ [banana (x)& y eat x]]
ii. ✔There is a banana that every baboon is eating
✔ $\exists x$ [banana (x) & $\forall y$ [baboon (y) $\rightarrow$ y eat x]]

(39) a. Ango n-o-ku-rɔ́r-a ***∅-βaa-na* (*β-ɔɔ-βy-ɔ́ɔsɛ)*** ha-yɔ?
b. *Ango n-o-ku-rɔ́r-a ***a-βaa-na* (*β-ɔɔ-βy-ɔ́ɔsɛ)*** ha-yɔ?
Q ?-2sg-IPFV-see-FV PPF-C2-child (any) C16-there
'Do you see any children there?'

(40) a. *Ango n-o-ku-rɔ́r -a a-βaa-na ha-yɔ?
b. Ango n-o-ku-rɔ́r-a ∅-βaa-na ha-yɔ?
Q 2sg-IPFV-see-FV PPF-C2-child C16-there
'Do you see any children here?'

Note that in example (39) it is optional to use the NPI *β-ɔɔ-βy-ɔ́ɔsɛ*' any' and for that matter there is no assertion of existence hence the overt PPF is illicit. In example (40) the overt PPF is licit in the context of assertion of the existence. This suggests that the material present in the sentence and/or context of use affects the construal of both overt and non-overt PPFs.

4.0 Conclusion

In this paper I have argued that PPFs in Nata are weak Ds that assert the existence of an entity and that they do not assert or presuppose familiarity or uniqueness/maximality. Nata PPFs/D (both overt and non-overt) share in common the property of assertion of existence of an entity. I have presented the data, which adduced that overt PPFs do not encode distal or proximal deictic force. I have argued against the hypothesis that nouns that appear with overt PPFs inherently assert or presuppose familiarity or uniqueness/maximality. I have shown that overt PPFs are definite if they are interpreted with definite material such as OMs, demonstratives, or possessives. With regard to non-overt PPFs, I have argued that their interpretation highly depends on context of use and/or the material available in the sentence just like their overt PPF counterparts. Although it is clear from the data I presented that both overt and non-overt PPFs are interpreted with respect to context- of-use and/or the presence of local elements in a sentence, a characteristic of weak D in Nata, the formal syntactic/semantic analysis of PPFs remains the object of further research. For instance, the question why non-overt PPFs are compatible with both assertion and non-assertion of existence or with both specific and non-specific is of theoretical importance. I have submitted that Nata arguments must be analyzed as DPs and PPFs as weak Ds. This step provides insights in the characterization and description of Nata PPFs. If this proposal is assumed then this work presents new evidence on the description and the interpretation of Bantu PPFs.

References

Ashton, E.O. et al. (1954). A Luganda Grammar . London: Longmans, Green and Co.
Bleek, W. H. I. (1869). A Comparative Grammar of South African languages. Trübner, London.
Chierchia, G. et al. (2008). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. MS, Harvard and MIT.
Déchaine, Rose-Marie & Mireille Tremblay. 2011. Deriving nominal reference. Paper presented at WECOL 2011, SFU.
Enç, M. (1991) "The semantics of specificity" Linguistic Inquiry 22: 1-25.
Fodor, J.D and I.A Sag (1982). "Referential and Quantificational indefinites". Linguistics and Philosophy 5, 355–398.
Frege, G. (1997 [1892]). On Sense and Reference. In P. Ludlow (Ed.), Readings in the Philosophy of Language, pp. 563–583. MIT Press.
Gambarage, J.J. (2012). The Pre-Prefix in Nata: An Interface Account. To appear in Proceedings of the 43rd Annual Conference on African Linguistics. University of Tulane, New Orleans.
Gillon, C. (2006). The semantics of determiners: domain restrictions in Skwxú7mesh . Ph.D. thesis, University of British Columbia, Vancouver, BC.
Givón, T. (1971). Studies in Chibemba and Bantu Grammar. Supplement 3 to Studies in African Linguistics.
Guillemin, D. (2007). The emergence of the determiner system in Mauritian Creole: A syntax semantics mapping. University of Queensland Working Papers in Linguistics. Vol.1. 1-19.

Guthrie, M. (1948). The Classification of Bantu Languages . London: Oxford University Press.
Heim, I [1982](2011). The Semantics of Definiteness and Indefinite Noun Phrases. PhD Dissertation. University of Massachusetts at Amhert.
Hyman, L. Katamba, F. & Walusimbi, L. 1987. Luganda and the strict layer hypothesis. Phonology (Yearbook) 4.87-108.
Ionin, T (2006) This is Definitely Specific: Specificity and Definiteness in Article Systems. Natural Language Semantics (2006) 14:175–234
Johannes, G.J (2007). The Ki-Nata Noun Structure. MA. Thesis. University of Dar es Salaam.
Krifka, Manfred (2003) Bare NPs : Kind-referring, Indefinites, Both, or Neither? In R. Young and Y. Zhou (eds), SALT XIII 1 80-203 , Ithaca, **,** : Cornell University.
Krifka, M and Zerbian, S (2008). "Quantification Across Bantu Languages". Center for General Linguistics, Berlin.
Lyons, C. (1999). Definiteness . Cambridge: Cambridge University Press.
Lyon, J (2011). The syntax of Okanagan determiner phrases. In Papers for the 46th Annual International Conference on Salish and Neighboring Languages , pp. ??.
Maho, J. (1999). A Comparative Study of Bantu Noun Classes. Cothoburgensin: Acta Universitatis Upsaliensis.
Matthewson, L. (1998). "Determiner Systems and Quantificational Strategies: Evidence from Salish". Holland Academic Graphics, The Hague.
Mattina, N. (2002). Determiner Phrases in Moses-Columbian Salish. International Journalof American Linguistics 72 (1), 97–134.
Mould, Martin. (1974). The syntax and semantics of the initial vowel in Luganda. In Proceedings of the 3rd Annual Conference on African Linguistics (E. Voeltz, ed.), pp. 223-229. Indiana University Press, Bloomington.
Progovac, L. (1993). Negative Polarity: Entailment and Binding. Linguistics and Philosophy. 16 (2):149-180.
Schneider-Zioga, P. and B. Handerson (2007). Anti-agreement, anti-locality and minimality: the syntax of dislocated subjects. Natural Language and Linguistic Theory, 25 (2), 403-446.
Reinhart, T. (1997), 'Quantifier Scope: How Labor is Divided between QR and Choice Functions,' Linguistics and Philosophy 20, 335–397.
Russell, B. (1998 [1905]). On denoting. In G. Ostertag (Ed.), Definite Descriptions: A Reader, Cambridge, MA, pp. 35–49. MIT Press.
Sharvy, R. (1980). A more general theory of definite descriptions. The Philosophical Review 89 , 607–624.
Visser, M (2008) Definiteness and specificity in the isiXhosa determiner Phrase. South Africa Journal of African Languages. Vol. 1, pp.11-29

Joash Johannes Gambarage
Joashj2010@gmail.com

Similarity and half-rhyme acceptability[1]

Kevin McMullin
UBC Department of Linguistics

The notion of similarity between sounds plays an important role in phonology and has been shown to be a contributing factor in many linguistic phenomena, including the frequencies of half-rhymes in several languages. Previous studies on half-rhymes have been done using various corpora, and have shown that there are inconsistencies in the data that cannot be explained using metrics of phonological similarity alone. The present study supplements the previous literature with an experimental study of English half-rhyme judgements. A measure of perceptual similarity is demonstrated to provide a better model of the observed data than two different measures of phonological similarity, and furthermore, phonological similarity adds nothing to the model. This is taken as evidence that language users have a detailed knowledge of the perceptual characteristics of the sounds of their language that is reflected in the relative acceptability as well as the relative frequency of different types of half-rhymes.

1 Introduction

1.1 Similarity

Some sounds of natural languages are more alike than others. This would perhaps not be surprising or interesting, except that this similarity plays an important role in many linguistic phenomena. A few of these include: speech errors (Stemberger 1991; Frisch 1996; Rose & King 2007), phoneme confusability in noise (Luce 1986; Woods *et al.* 2010), short-term memory confusion of phonemes (Wicklegren 1966), the nature of phonotactic constraints (Frisch *et al*. 2004), ease of learning phonotactic dependencies (Moreton 2012), and, of particular interest here, half-rhyme frequencies (Steriade 2003; Kawahara 2007). Determining the appropriate quantitative measure of similarity has been a topic of interest in the literature, and recently there has been much discussion surrounding the relevance of metrics for different types of similarity (Gallagher & Graff 2012) and how they could potentially interact.

There are at least four main types of similarity between phonemes that have been proposed in the literature, each with its own set of measurements or calculations: acoustic similarity, articulatory similarity, phonological similarity, and perceptual similarity. I will provide a brief description of each of these below, with particular attention paid to phonological and perceptual similarity, as they are most relevant to the current study.

Measures of acoustic similarity are obtained by comparing the numbers of a raw acoustic measure between sounds. This is most easily done for a small set of sounds that share some phonetic characteristic with an easily quantified acoustic measure (e.g. VOT, vowel formant frequencies, pitch).

Articulatory similarity refers to the extent that two sounds use similar articulators or have similar muscle activations when produced. The most extensive study of this type of similarity is found in Mielke (2012), which investigates the relationship between articulatory measures of similarity between sounds (obtained by comparing oral airflow, nasal airflow, vocal fold contact, and larynx height) and how these sounds pattern typologically. This type of similarity could potentially be quite useful in future research, but determining the best measures to take and obtaining these measures each pose their own set of problems.

Measures of phonological similarity are obtained by some phonologically informed calculation that determines how similar two phonemes are. Many of these measures are based on the number of

[1] The research reported here was supported by the Social Sciences and Humanities Research Council of Canada grant #F10-04808, awarded to Molly Babel.

shared distinctive features (Bailey & Hahn 2005; Rose & King 2007). Another popular measure of phonological similarity is Frisch's natural class metric (Frisch 1996; Frisch *et al.* 2004), which is calculated with the following formula:

$$(1)\quad \textit{Phonological Similarity} = \frac{\textit{Shared natural classes}}{\textit{Shared natural classes} + \textit{Unshared natural classes}}$$

This measure has been shown to be especially good at modeling English speech errors (Frisch 1996), as well as a gradient OCP-Place constraint in Arabic (Frisch *et al.* 2004), but performs poorly for co-occurrence restrictions in Muna (Coetzee & Pater 2008) and Gitksan (Brown & Hansson 2008). For this reason, the present study uses two separate measures of phonological similarity – one based only on features, and one based on Frisch's natural class metric.

Perceptual similarity is a measure of how similar two phonemes are perceived to be. One important note about metrics for perceptual similarity is that they must be obtained indirectly through experimental studies. This can be done with confusion-in-noise studies, forced-choice tasks, or by having subjects rate similarity on a scale. One common argument is that perceptual similarity plays an important role in the shaping of phonological constraints (e.g. the P-Map hypothesis from Steriade 2001; 2003), and more recently that it plays an important role in feature emergence (Mielke 2008) as well as emergent constraint rankings (Johnsen 2012). Throughout this paper I will use a measure of perceptual similarity that is adapted from Woods *et al*. (2010), which carried out an extensive confusion-in-noise experiment that is described in more detail in section 3.1.

1.2 Rhymes and half-rhymes

Rhyming is a tradition in many languages, pervasive in poems, songs, puns, and language games. While the exact definition of a rhyme differs slightly across languages, it seems that the common goal of rhyming is to make the ends of words sound the same. In Japanese, for example, the traditional definition of a rhyme requires that the ends of two rhyming lines have all the same vowels, but the onset consonants need not match (Kawahara 2007). The traditional definition of a perfect rhyme in English is different in that it requires that two words have an identical nucleus of primary stress, and that all the following segments are also identical (Rickert 1979). Some examples of perfect rhymes in English include {pack, stack} and {surrender, defender}. Note that the conditions of this definition are not always met for two words found in a rhyming pair. Consider, for example, the line from a popular English children's song *This Old Man*: "With a knick-knack paddy-whack give a dog a **bone** / This old man came rolling **home**". The rhyming pair here {bone, home} shares the nucleus of primary stress, but there is a mismatch in the word-final coda consonants. Pairs like this that deviate from the traditional definition of a rhyme are referred to as *half-rhymes* (sometimes imperfect rhymes). English speakers seem to share the intuition that some half-rhymes are more tolerable than others. For example, while {time; line} is a relatively good half-rhyme, {time; like} is not as good. The usual assumption is that the mismatched sounds in the better half-rhyme pairs must be more similar to each other. Half-rhymes can also include pairs that differ in their vowels, their number of coda consonants, or their number of syllables, but this study is primarily concerned with half-rhymes like the above examples, where there is a mismatch of two single coda consonants.

1.3 Half-rhymes and similarity in the literature

Half-rhymes have been used as a tool to investigate historical sound changes, particularly in English (Wyld 1923; Sonderegger 2011), as well as the notion of similarity (Zwicky 1976; Steriade 2003; Kawahara 2007). For the latter, this has been done exclusively through corpus studies, which have been successful in showing that indeed some types of half-rhymes are more common than others and that these frequencies generally coincide with some measure of similarity, but various problems are left unresolved.

Zwicky (1976) studied half-rhymes in English rock lyrics and makes some observations that have not yet been accounted for. These include what he calls the *gap problem* and the *long-distance problem*.

The gap problem is that some pairs of phonemes that are quite similar in terms of features are not found very often in half-rhymes. His example is that voiced-voiceless pairs, such as {b,p}, which differ only in the voicing feature, are rarely paired together as a half-rhyme. He cites two exceptions to this *gap*, namely the {t,d} and {s,z} pairs. The long-distance problem is just the opposite. Some phoneme pairs that differ in at least two features, and are therefore farther apart with respect to similarity, such as {b,z} are commonly found in half-rhyme pairs. These two problems can be summarized as follows: some consonant pairs differing only by one feature are not found very often, while some consonant pairs differing by several features are quite common. The same problems apply even when using the more complex measure of phonological similarity that is based on natural classes, which is given in (1). The present study gives evidence for a good solution to these problems if we take perceptual similarity as the best measure of a good half-rhyme.

Kawahara (2007) examines half-rhymes in Japanese rap lyrics. I noted above that the definition of a Japanese rhyme does not require the consonants to be identical. However, Kawahara finds that two consonants are more likely to be in correspondence in a rhyming pair if they are identical or quite similar to each other. He confirms this by applying Frisch's natural class metric of similarity (Frisch 1996) to the data and finding a positive correlation. He notes, however, that several pairs were not acting as expected, and attributes this to their acoustic similarity. For example, the {kʲ, tʃ} pairs, though they share few natural classes, are highly overrepresented in the corpus (this is another example of the long-distance problem that Zwicky (1976) describes). He argues that this is a result of their acoustic similarity and that the source of this must be that the composers of Japanese rap lyrics have a detailed knowledge of the acoustics of Japanese consonants, including acoustic similarities that cannot be captured by any phonological categorization of the sounds. However, he also admits that his study is incomplete as we would need some measure of acoustic or perceptual similarity of Japanese consonants to compare his results to, and he suggests that a comparison with a confusion matrix would be useful for future research.

McMullin (2009) follows up on Kawahara's study by examining a corpus of half-rhymes within the lyrics of the Red Hot Chili Peppers. The results show that coda consonant pairs sharing voicing, continuancy, or nasal features are overrepresented in the data, and that pairs differing in these features are underrepresented. Conversely, consonant pairs that share place of articulation (labial, alveolar, or velar), are all underrepresented. This asymmetry is taken as evidence that it is more than just the phonological similarity of the consonants that is underlying half-rhyme acceptability. Consider a pair of words whose coda consonants share a voicing feature but not place, such as {ca**b**, la**d**}, in comparison with a pair whose coda consonants differ only in voicing, such as {ca**b**, la**p**}. The length of the preceding vowel is approximately the same for the {b,d} pair, but the length of the vowel preceding a [p] is much shorter. As a result, the VC rhyme of {b,d} forms a better half-rhyme than {b,p}, despite any acoustic consequences of changing the place of articulation. It seems that a measure of similarity needs to also take the syllabic position of the sound into consideration. Phonological similarity remains the same regardless of position in the word or syllable, whereas perceptual similarity can take this into account by taking scope over the syllable nucleus as well as the coda.

The biggest problem with the corpus studies described above is that they show systematic deviations from what metrics of phonological similarity predict. These have been pointed out and discussed, with a potential solution relying on perceptual similarity, but this explanation has never been tested. The present study uses an experimental study of English half-rhyme judgements to investigate the hypothesis that perceptual similarity will provide a better model than phonological similarity. Results indicate that perceptual similarity is indeed the best measure for modeling half-rhyme acceptability and furthermore that including a measure of phonological similarity in the model does not improve it.

2 Methods

2.1 Subjects

Ten subjects (7 female, 3 male, mean age of 24 years) participated in the study. All were native English speakers and reported no speech, language or hearing problems. Subjects were compensated $10 each for a task that took approximately 50 minutes to complete.

2.2 Stimuli

All words used in the stimuli set were extracted from the SUBTLEXus database (Brysbaert & New 2009), using a script which extracted all monomorphemic CVC words that had one of the codas [p,b,t,d,m,n] and one of the vowels [i,ɪ,e,ɛ,æ,ʌ,u,o,ɔ]. The coda consonants were chosen so that there were minimal pairs for each of place (labial vs. alveolar), voicing, and nasal features. In order to allow for many possible stimuli combinations, onset consonants were not controlled. Finally, the 25% least and most frequent of these words were excluded to help control for any frequency bias the subjects might have. Upon completion of the list, 90 words were chosen such that there were 15 with each of the six coda consonants (refer to Appendix 1 for a complete list of the 90 words). An adult male native speaker of North American English then recorded the stimuli in a soundproof booth. He was unaware that the words would be used in a study of half-rhymes, and each word was presented to him in a random order.

2.3 Design and procedure

Stimuli were presented to the subjects on headphones in a quiet room. On each trial, subjects were presented with two pairs of CVC words and picked which pair of words formed a better rhyme by pressing a button. For each trial, the vowel was the same for all words, and the first word of the pairs was identical. For example, one trial was {robe, load ; robe, foam}, which compares the relative acceptability of {b,d} and {b,m} pairs in half-rhymes. There was a 100ms interval between the two words within each pair and a 500ms interval between the two pairs. Subjects were given a maximum of three seconds after the onset of the final word to register a response and were presented with their reaction time after each trial to help maintain interest. In total there were three sets of pairs for each of the 60 possible comparisons, counterbalanced for order, for a total of 360 trials per subject. Each subject was given a short break halfway through the experiment. For the remainder of this paper, each trial is referenced by the coda consonants that were presented in the two pairs. Using the same example as above, a trial that compares the relative acceptability of {b,d} and {b,m} half-rhymes is referred to as a “bdbm” trial.

3 Results and analysis

3.1 Measures of similarity

In order to compare the results of this experiment to phonological similarity, I will use one measure that is based on natural classes (as in Frisch 1996; Frisch *et al.* 2004) and another measure that is based only on features. It is important to note that these metrics, along with all metrics of phonological similarity, are dependent on a particular analysis of the phonological features and natural classes of English. The relevant values for natural class similarity are shown in Table 1 below while those for feature similarity are shown in Table 2 below.

Table 1 – Natural class similarity as given in Frisch (1996: p.45)

Consonant	d	m	n	p	t
b	0.28	0.39	0.12	0.40	0.14
d		0.11	0.38	0.14	0.39
m			0.26	0.19	0.06
n				0.06	0.19
p					0.30

Table 2 – Number of common feature values for [voice], [labial], [nasal], and [cons][2]

Consonant	d	m	n	p	t
b	3	3	2	3	2
d		2	3	2	3
m			3	2	1
n				1	2
p					3

To obtain a measure of perceptual similarity, I will use data from Woods *et al.* (2010). The authors carried out an extensive confusion-in-noise experiment that included over 34000 trials, with CVC stimuli presented in different levels of noise, and allowed for responses that included any combination of phonotactically permissible English CVC syllables, both words and non-words. As a result, we can focus our attention on the confusions of coda consonants. The authors report the raw numbers of responses and each consonant was presented an equal number of times in their study, so there is no need to normalize the numbers. The confusions relevant to the present study are shown in Table 3 below.

Table 3 – Number of confusions as given in Woods *et al.* (2010: p.1615)

Consonant	d	m	n	p	t
b	197	46	28	172	16
d		19	53	28	49
m			341	15	12
n				10	15
p					213

3.2 A comparison in percentages

In this section, the results are presented in terms of the number of times that the first pair of a trial was chosen, pooled across subjects, and expressed as a percentage. In Figures 1-3, these observed percentages are plotted on the vertical axis and, for each type of similarity, are compared to the expected percentages of first pair choices, plotted on the horizontal axis. The expected percentages were calculated by the formula given below:

$$\text{(2)}\quad \textit{Expected \%} = \frac{\textit{1}^{st}\ \textit{pair similarity}}{\textit{1}^{st}\ \textit{pair similarity} + \textit{2}^{nd}\ \textit{pair similarity}} \times \textit{100}$$

As an example, the expected percentage of {b,d} choices in the bdbm trials is 41.8% for natural class similarity, 50% for feature similarity, and 81.1% for perceptual similarity.

[2] The [cons] feature is included to avoid the complications of dividing by zero when comparing two values of feature similarity.

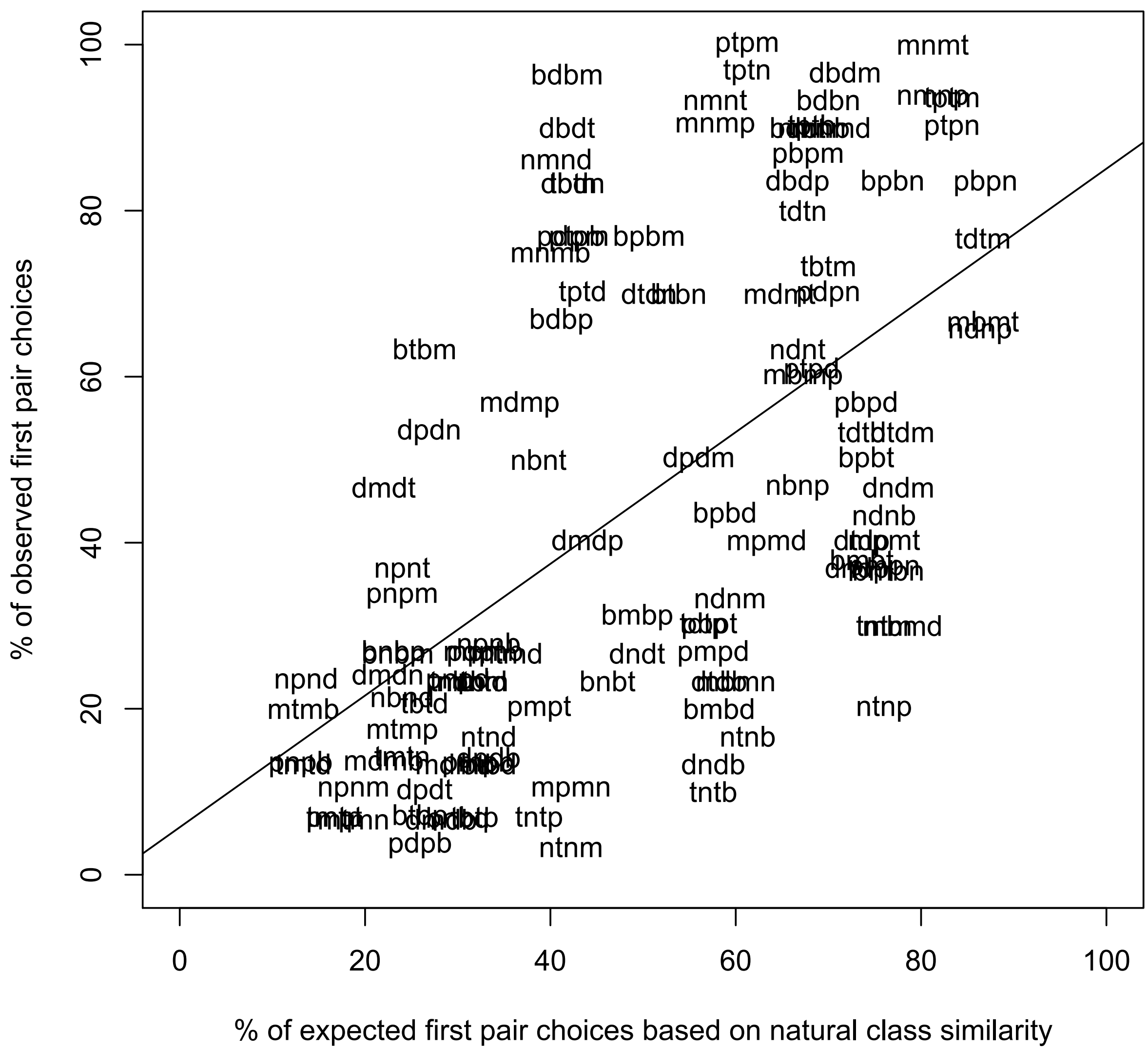


Figure 1 – Observed vs. expected (based on natural class similarity) choices of the first pair for each comparison. The solid black line represents the line of best fit.

In the above figure, the expected percentages are significantly correlated with the observed data (Pearson's r=0.566, $p<0.001$), showing that the first pair is more likely to be chosen if it has a higher value of natural class similarity compared to the second pair.

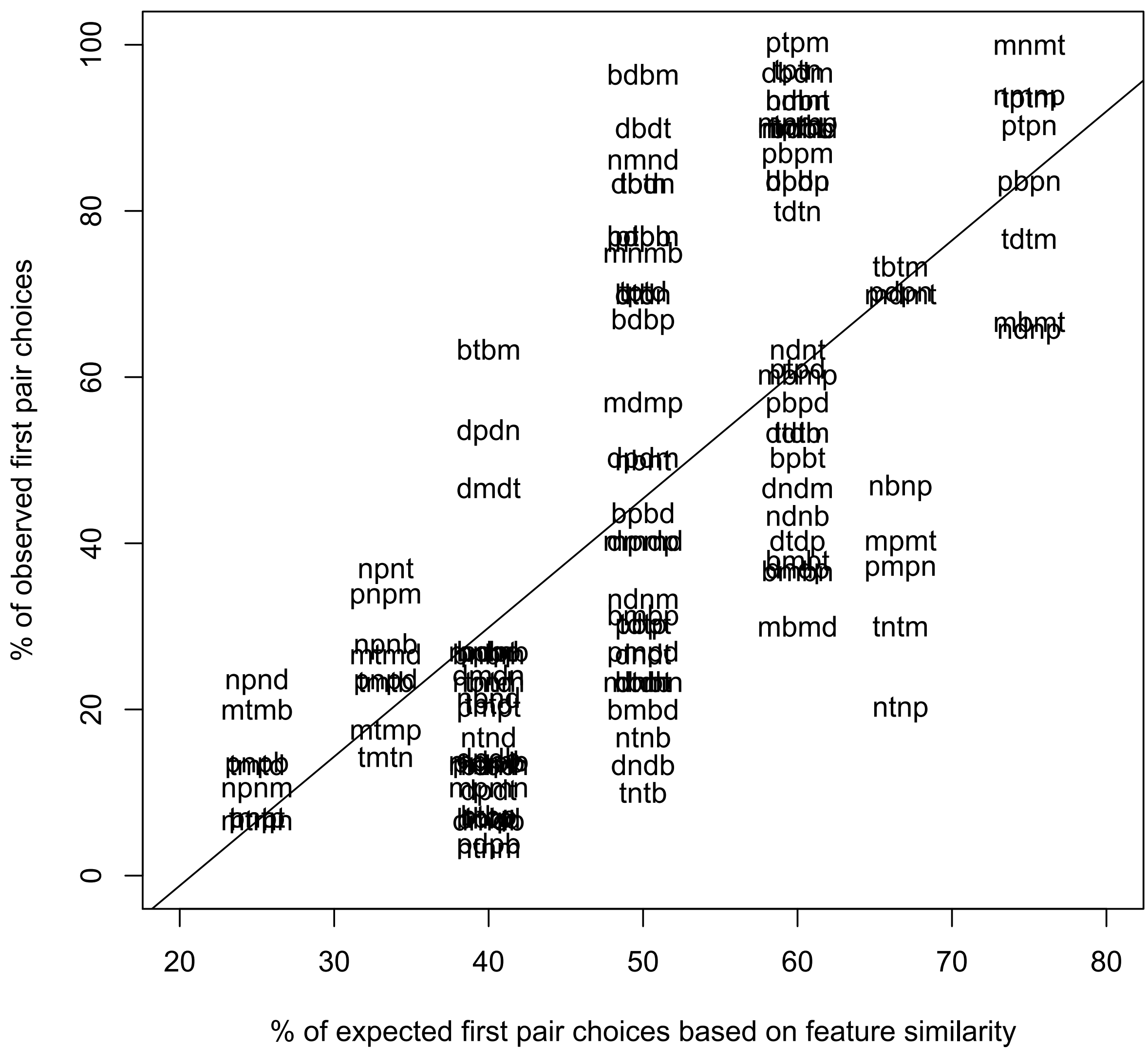


Figure 2 – Observed vs. expected (based on feature similarity) choices of the first pair for each comparison. The solid black line represents the line of best fit.

In the above figure, the expected percentages are significantly correlated with the observed data (Pearson's r=0.678, $p<0.001$), showing that the first pair is more likely to be chosen if it has a higher value of feature similarity compared to the second pair.

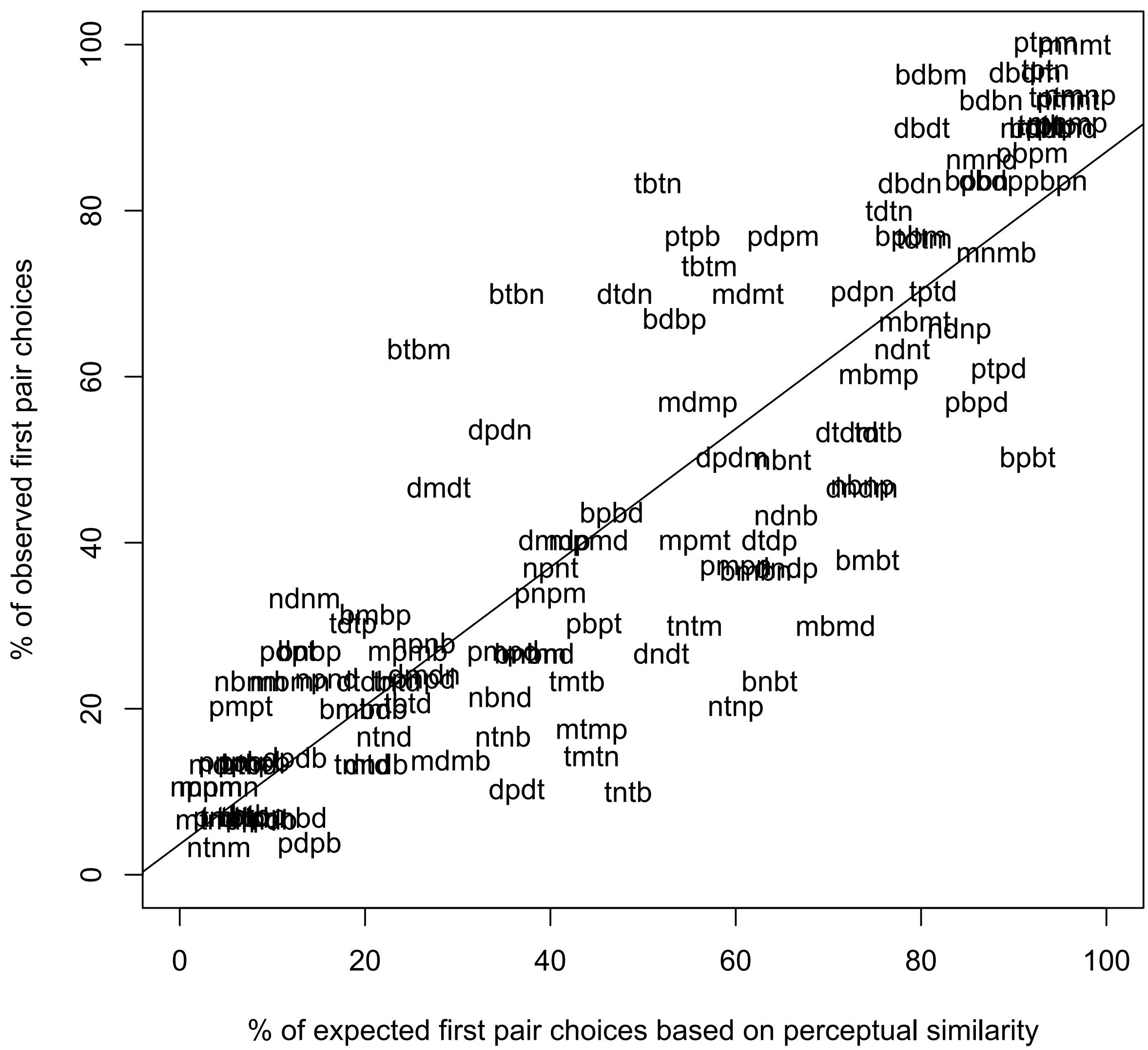


Figure 3 – Observed vs. expected (based on perceptual similarity) choices of the first pair for each comparison. The solid black line represents the line of best fit.

In the above figure, the expected percentages are significantly correlated with the observed data (Pearson's r=0.863, p<0.001), showing that the first pair is more likely to be chosen if it has a higher value of perceptual similarity compared to the second pair.

The goal of this section was to provide an intuitive comparison of the subjects' responses and the three measures of similarity. The results can be summarized as follows: of the two measures of phonological similarity, the expected percentages based on feature similarity are more highly correlated with the observed responses than those based on the natural class metric or similarity, while the expected percentages based on perceptual similarity are even more highly correlated. However, these numbers are simply given in raw percentages (bounded between 0 and 100), obtained by pooling across subjects that may have different tendencies in their responses. In the following section, I will analyze the data more completely by building up and comparing several mixed-effects logistic regression models.

3.3 Mixed-effects logistic regression models of the data

Each trial in this study required the subject to choose which of two pairs was the better half-rhyme: the first pair or the second. One way of addressing this categorical type of data is with a logistic regression model, which uses a set of independent variables to predict the log odds of choosing one of two categorical values (see Baayen 2008; Jaeger 2008 for a description of these models and their advantages). This section presents a summary of several logistic regression models that predict the log odds of choosing the first pair of each trial, based on a comparison of the relative similarities for the two pairs. An advantage of mixed-effects models is that variables that were not controlled in the design can be built into the model with random intercepts and slopes. Barr *et al.* (in press) argues for a maximally rich random-effects structure that is justified by the design. The models that follow assign random by-subject intercepts and random slopes for each subject and for each vowel. All models below were created with the *glmer* function from the *lme4* package (Bates *et al.* 2012) implemented in R (R Core Team 2012).

Natural class similarity – In order to use natural class similarity as a predictor in the logistic regression models, I will need one number that compares the relative similarity of the two pairs. The calculation used to obtain this comparison is given below in (3):

$$(3)\ \ \textit{Natural class comparison} = \textit{ln}\left(\frac{\textit{natural class similarity of } 1^{st} \textit{ pair}}{\textit{natural class similarity of } 2^{nd} \textit{ pair}}\right)$$

This comparison creates an unbounded continuous measure centered around zero. By taking the natural log, it is no longer subject to the problems created by using a strict ratio that concentrates half of its values between 0 and 1. As this number increases, so too does the relative similarity of the first pair compared to the second pair. If the natural class comparison is greater than zero, then the first pair is more similar than the second pair. If it equals zero, then the two pairs are equally similar. If it is less than zero, then the second pair is more similar than the first pair. A summary of the model that uses this comparison of natural class similarity is given in Table 4 below.

Table 4
A summary of the fixed effects of mixed-effects logistic regression with natural class similarity (N = 3573, # of Subjects = 10, AIC = 4481)

Coefficient	Estimate	SE	Pr(>|z|)
Intercept	–0.2135	0.0855	0.013*
Natural class comparison	0.7791	0.0623	<0.001*

Both the intercept and the Natural class comparison make significant contributions in the above model. The coefficient of the intercept is negative, which means that overall, subjects are biased towards choosing the second pair (in the raw data, they did so approximately 55% of the time[3]). The estimate for the contribution of the Natural class comparison is a positive value, indicating that when the Natural class comparison increases (which means that the first pair is becoming more similar compared to the second pair), so too do the odds that the first pair is chosen.

Feature similarity – To build a model using feature similarity, I will need a calculation similar to that used in (3), but which gives one number that compares the number of shared features between the two pairs. The formula I used is shown in (4) with the resulting model given in Table 5:

[3] This is not an unexpected result given the nature of the task. It may be that this is some sort of recency effect, in which subjects are more likely to pick the pair that they most recently heard. Another possibility is that, since none of the options are perfect rhymes, subjects may be prone to make a decision that the first pair is a relatively bad rhyme even before hearing the second pair, causing more second pair choices. A full analysis of this effect is outside the scope of this study.

$$(4)\ \ Feature\ comparison = ln\left(\frac{\#\ of\ common\ features\ of\ 1^{st}\ \ pair}{\#\ of\ common\ features\ of\ 2^{nd}\ \ pair}\right)$$

Table 5
A summary of the fixed effects of mixed-effects logistic regression with feature similarity (*N* = 3573, # of Subjects = 10, AIC = 4281)

Coefficient	**Estimate**	**SE**	**Pr(>|z|)**
Intercept	–0.2270	0.0906	0.012*
Feature comparison	1.7688	0.1216	<0.001*

The above model has a negative intercept that makes a significant contribution to the model, as well as a positive estimate for the Feature comparison that also makes a significant contribution to the model. Note that as the relative similarity, with respect to features, of the first pair increases compared to the second pair, so does the Feature comparison. As a result, this positive coefficient shows that subjects are more likely to choose the first pair in a trial as its feature similarity increases compared to the second pair.

Perceptual similarity – With the same procedure as above, the calculation for a comparison of perceptual similarity is given in (5) below (based on Woods *et al.* 2010) with a summary of the model in Table 6:

$$(5)\ \ Perceptual\ comparison = ln\left(\frac{\#\ of\ confusions\ for\ 1^{st}\ \ pair}{\#\ of\ confusions\ for\ 2^{nd}\ \ pair}\right)$$

Table 6
A summary of the fixed effects of mixed-effects logistic regression with perceptual similarity (*N* = 3573, # of Subjects = 10, AIC = 3799)

Coefficient	**Estimate**	**SE**	**Pr(>|z|)**
Intercept	–0.2839	0.1116	0.011*
Perceptual comparison	0.8179	0.0793	<0.001*

This model has a negative intercept that is significant, which once again shows that subjects are biased towards choosing the second pair. The Perceptual comparison also makes a significant contribution to the model, with a positive coefficient, showing that as the first pair becomes more perceptually similar compared to the second pair, the odds of choosing the first pair increase.

Model comparison – Tables 4, 5 and 6 show that both measures of phonological similarity as well as the measure of perceptual similarity are significant predictors of the subjects' responses (this is in line with section 3.1 above, in which the observed percentages of first-pair choices were significantly correlated with the expected percentages for all three measures of similarity). However, the crucial question is whether one model, and hence predictor, is better than the others. Note that a step-wise comparison of these models is not possible since they are not nested, and they do not share any predictor variables. An alternative method for comparing models, which I will adopt here, is to use the *Akaike information criterion* (AIC) (Akaike 1974; 1980) to compare models. This number, which is reported for each model in Tables 4, 5, and 6 above, is an estimation of the information lost when using that model to predict the observed data. The model with the lowest AIC gives the best fit to the data.

An additional property of the AIC, since it gives only an estimation of the information lost, is that we can calculate the odds that one model is better than another. This is known as the *Akaike weight*. The formula for doing this, given in (6), gives the odds that model *i* minimizes the information lost, as compared to the model with the minimum AIC value.

(6) $Akaike\ weight = w_i = exp\left(\frac{AIC_{min} - AIC_i}{2}\right)$

Using the AIC values from Tables 4 and 5 (AIC = 4481 for natural class similarity; AIC = 4281 for feature similarity), we can calculate that the model that uses natural class similarity has a probability of less than 10^{-44} (effectively 0) of minimizing the information lost, as compared to the model that counts up the common features for each pair. We can conclude that feature similarity is the better phonological measure for modeling the observed percentages in half-rhyme judgments. However, a comparison of this model and the model using the perceptual measure of similarity as a predictor (AIC = 3799) shows that there is a probability of less than 10^{-105} that the model that uses feature similarity minimizes the information lost compared to the model that uses perceptual similarity. We can conclude that perceptual similarity is the best single measure of the three for modeling subjects' judgements of half-rhyme acceptability.

3.3 Including more than one type of similarity in the model

While the analysis provided in section 3.2 showed that perceptual similarity is the better measure for modeling the relative acceptability of English half-rhyme pairs, it should still be considered whether phonological similarity can make any contribution at all. In this section I will build up another mixed-effects logistic regression model that uses both the perceptual measure of similarity and the measure of phonological similarity based on features (the better of the two phonological measures) as predictors to show that adding phonological similarity to the model makes no significant contribution to the model.

Note, however, that *multi-collinearity* can be a problematic issue in this type of analysis. That is, the measures in (4) and (5), which give comparisons of the feature and perceptual similarities for each pair, are significantly correlated with each other (Pearson's r=0.779, p<0.001). This can lead to several problems when including them both in a model (Baayen 2008), the most troublesome of which is that the estimates for the effect size of the two collinear measures can vary unpredictably, with neither coming out as significant. However, feature similarity has systematic differences from perceptual similarity, so it should be possible to separate them and instead build those differences into the model. To do this, I first created a linear regression model of the feature similarity measures using perceptual similarity. I then took the residuals of this model. These residuals are completely uncorrelated with perceptual similarity, and are included in the model in Table 7. [4]

Table 7
A summary of the fixed effects of mixed-effects logistic regression with perceptual similarity and the residuals from feature similarity (N = 3573, # of Subjects = 10, AIC = 3801)

Coefficient	Estimate	SE	Pr(>|z|)
Intercept	–0.2845	0.1124	0.011*
Perceptual comparison	0.8264	0.0819	<0.001*
Feature residuals	–0.1663	0.1605	0.300

Note that while the Intercept and the Perceptual comparison still provide significant contributions to the model, the additional measure of feature similarity (as it systematically differs from perceptual similarity) does not come out as a significant predictor of the data. Furthermore, the above model (AIC = 3801) actually performs worse than the model that includes perceptual similarity alone (AIC = 3799). Using the calculation of Akaike weight given in (6), the odds are only 0.36 that this more complex model minimizes the information lost, when compared with the simpler model that I based only on perceptual similarity.

[4] It turns out that a model that uses the measure of feature similarity instead of the residuals results in the same AIC value, as does a model that uses feature similarity as the main measure and includes the residuals of perceptual similarity (though in the latter case, both predictors make significant contributions, as would be expected).

4 Discussion

The main goal of this study was to determine which measure of similarity was best for predicting half-rhyme acceptability. The analysis presented above provides evidence that perceptual similarity by itself is the best measure and that a combination with phonological similarity does not give any improvement. With this in mind, we can address the problems specific to the connection between similarity and half-rhymes. We can also explore the implications, beyond half-rhymes, of the idea that language users have a detailed knowledge of the perceptual characteristics of the sounds in their language, and the possibility that there is more than one notion of similarity that they can access.

4.1 Half-rhymes and similarity

Recall that Zwicky (1976) noticed two problems with the idea that phonological similarity is behind half-rhyme frequencies. The gap problem is that some pairs that differ by only one feature are uncommon in half-rhymes. As an example, he noted that pairs differing only in voicing were quite rare. This underrepresentation of voiced-voiceless pairs has been replicated in this experiment. On trials in which voiced-voiceless pairs were presented against labial-coronal pairs (e.g. bdbp, tdtp), subjects preferred the mismatch in place of articulation in 72.2% of the trials. I attribute this to the consequence that a change in voicing has on the preceding vowel length in English. Phonological similarity cannot account for any acoustic consequences outside of the consonant itself, but perceptual similarity can take scope over the preceding vowel allowing it to have an influence in the acceptability of a half-rhyme pair that, phonologically, differs only in its final consonant. With respect to Zwicky's exceptions for the gap problem, the {t,d} and {s,z} pairs, note that since his study was based only on raw counts of half-rhymes, it is not surprising that are the two common pairs, as they include phonemes that are found in English inflectional morphology (past tense, plural, 1st person singular, possessives). If he had performed an analysis that took their overall frequency into account, I expect that these pairs would also be underrepresented. Although this experiment does not offer any obvious explanation of the long-distance problem, that some consonant pairs differing by several distinctive features are found quite often in half-rhymes, we can use this same line of thinking to address the example given in Kawahara's (2007) study of Japanese half-rhymes – the {kʲ, tʃ} pair is highly overrepresented. This example is particularly telling because it includes two consonants that are phonologically dissimilar, but that are quite similar acoustically (and likely perceptually). If language users had no notion of similarity outside of phonological representations when composing half-rhymes, pairs like this should be uncommon. However, if they have a knowledge of the perceptual and acoustic characteristics of the sounds in their language, finding that pairs like this are relatively common is not at all surprising.

In the study presented in this paper, all words in the half-rhyme pairs were one-syllable CVC words with mismatches in the coda consonants. However, half-rhymes are not always of this form. A useful follow-up would be to have subjects perform the same task, but instead give judgements on stress-initial CVCV half-rhyme pairs that differ in the consonants (e.g. {baby, lady} as a {b,d} pair). If subjects are indeed using perceptual similarity to inform their acceptability judgements, then a measure of the perceptual similarity between consonants in onset position should be the best way to model the responses. Note that these numbers will systematically differ from those for consonants in coda position, since the acoustic effects on the surrounding vowels will not be as strong. If we find that English speakers judge the relative acceptability of half-rhymes differently when the same consonants are in a different positions, then we will have even stronger evidence that they are using perceptual similarity.

One potential criticism of the conclusions drawn from this study is that in the experimental procedure subjects actually heard the pairs of half-rhymes that they were asked to judge, and that all of these words were produced by the same speaker. Consequently, they have direct access to the acoustic signal and can use that information to inform their choice, rather than some implicit knowledge of the perceptual similarity of the half-rhyme pairs. Therefore, it is important that the results from this study are in line with results from corpus studies of half-rhymes, which include rhymes that were actually composed, not forced on a listener. However, to help avoid this confound, future studies of half-rhyme acceptability will include items produced by several speakers to determine whether half-rhyme pairs, even when produced by acoustically dissimilar voices, is still judged in the same way.

4.2 Further implications

While Frisch's natural class measure of similarity has proven to be useful in approximating the results of several processes that are attributed to similarity, we need to consider what the goal is for each of these processes. For example, for half-rhymes, the goal seems to be to make the end of a word or phrase sound as close to another one as possible. It is a perceptual goal, and I have shown with this study that a measure of perceptual similarity is far better at modeling the results of a half-rhyme judgment task than a measure of phonological similarity. However, it seems unlikely that perceptual similarity is underlying all of the phenomena listed in section 1.1, such OCP constraints, speech errors and so on. A more reasonable approach is that each of these phenomena could be paired with different types of similarity, and possibly a combination of two or more types of similarity. Consider a gradient OCP-place constraint, such as that found in Arabic (McCarthy 1986; Frisch et al. 2004), which essentially prohibits two consonants within a root from having the same place of articulation. There are multiple possible goals for such a constraint. It could be to ease the burden put on a speaker from having to make two sounds at the same place of articulation. If this is the case, we might expect that some measure of articulatory similarity would be best at approximating the data. It could be, however, that having different places of articulation in a word root makes it easier for a listener to perceive the sounds correctly. If this is the case, we would expect that a measure of acoustic or perceptual similarity for Arabic consonants would best describe the constraint. Allowing for multiple types of similarity to play a role in phonology leads to some interesting and unexplored questions. Are all of these types of similarity stored separately, or as one representation that has multiple dimensions? Can certain measures be completely ignored for some behaviors but not others?

5 Conclusion

Using a forced choice task, I have shown that some half-rhyme pairs are systematically judged to be better than others. Beyond this, a measure of perceptual similarity (based on Woods et al. 2010) was shown to be best at modeling the subjects' responses and that adding a measure of phonological similarity does not improve the model. This is just one application of the notion of similarity, and we should not be treating all of the linguistic phenomena attributed to similarity in the same way. With respect to half-rhymes, further perception studies that use multiple speakers and consonants in different syllabic positions will help to determine whether English speakers are using an internal knowledge of similarity to inform their judgements of half-rhymes. With respect to similarity in general, it would be useful to examine the other linguistic phenomena, such as OCP-place constraints to determine whether a different type of similarity or combination of these measures provides us with the best model. One then can we begin to address some of the bigger questions about how similarity is represented and how it should be incorporated into linguistic theories.

Appendix 1 – Stimuli

	Coda					
Vowel	b	d	m	n	p	t
i		seed	beam	teen	sheep leap	wheat
ɪ	rib	mid	limb	fin	dip	kit
e	babe	raid shade	fame shame	cane chain	shape tape	bait gate hate
ɛ	web	shed	gem	hen	rep	jet
æ	cab tab lab	pad tad	ram sham	fan ban	gap	bat chat
ʌ	pub rub	dud thud	hum gum	nun bun	cup pup	putt gut
u	tube boob	rude dude	zoom tomb	dune noon	soup hoop	loot boot
o	lobe robe	load code	foam dome	moan cone	cope mope	vote goat
ɔ	sob job knob	nod pod	bomb calm	con don	mop top	dot

References

Akaike, Hirotugu. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–723.

Akaike, Hirotugu. (1980). Likelihood and the Bayes procedure. In Bernardo, J.M. (Ed.) *Bayesian Statistics*. University Press, Valencia: 143–166.

Baayen, R. Harold. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge University Press.

Bailey, Todd M. and Ulrike Hahn. (2005). Phoneme similarity and confusability. *Journal of Memory and Language* 44: 568–591.

Barr, D.J., R. Levy, C. Scheepers, and H.J. Tily. (in press). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*.

Bates, Douglas, Martin Maechler, and Ben Bolker. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. http://CRAN.R-project.org/package=lme4.

Brown, Jason and Gunnar Ólafur Hansson. (2008). Lexical consonant cooccurrence patterns in Gitksan (Tsimshianic). Poster presented at the annual meeting of the Linguistic Society of America, Chicago, January 2008.

Brysbaert, M. and B. New. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behaviour Research Methods* 41: 990–997.

Coetzee, Andries W. and Joe Pater. (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistics Theory* 26: 289–337.

Frisch, Stefan. (1996). *Similarity and Frequency in Phonology*. PhD Thesis, Northwestern University. Evanston, Illinois.

Frisch, Stefan, Janet Pierrehumbert, and Michael Broe. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22: 179-228.

Gallagher, Gillian and Peter Graff. (2012). The role of similarity in phonology. *Lingua* 122: 107–111.

Jaeger, T. Florian. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towardsLogit Mixed Models. *Journal of Memory and Language,* 59: 434–446.

Johnsen, Sverre S. (2012). From perception to phonology: The emergence of perceptually motivated constraint rankings. *Lingua* 122: 125–143.

Kawahara, Shigeto. (2007). Half rhymes in Japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics*, 16: 113-144.

Luce, P.A. (1986). *Neighborhoods of words in the mental lexicon.* Ph.D. Thesis, University of Indiana. Bloomington, Indiana.

McCarthy, John. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17: 207–263.

McMullin, Kevin. (2009). Phonological similarity and half-rhymes in Red Hot Chili Peppers' lyrics. Unpublished manuscript. University of British Columbia.

Mielke, Jeff. (2008). *The Emergence of Distinctive Features*. Oxford University Press, Oxford.

Mielke, Jeff. (2012). A Phonetically based metric of sound similarity. *Lingua* 122: 145–163.

Moreton, Elliott. (2012). Inter- and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language* 67: 165–183.

R Core Team. (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rickert, William. (1979). Rhyme terms. *Style* 12: 35–43.

Rose, Sharon and Lisa King. (2007). Speech error elicitation and co-occurrence restrictions in two Ethiopian Semitic languages. *Language and Speech* 50: 451–504.

Sonderegger, Morgan. (2011). Applications of graph theory to an English rhyming corpus. *Computer Speech and Language* 25: 655–678.

Stemberger, Joseph. (1991). Radical underspecification in language production. *Phonology* 8: 73–112.

Steriade, Donca. (2001). Directional assymetries in place assimilation. A perceptual account. In: Hume, E., Johnson, K (Eds.). *The Role of Speech Perception in Phonology*. Academic Press, San Diego, CA: 219–250.

Steriade, Donca. (2003). Knowledge of similarity and narrow lexical override. *Proceedings of Berkeley Linguistics Society* 9: 583-598.

Wicklegren, W.A. (1966). Distinctive Features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America* 39: 388-398.

Woods, David, E. WilliamYund, Timothy Herron, and Matthew Ua Cruadhlaoich. (2010). Consonant identification in consonant-vowel-consonant syllables in speech-spectrum noise. *Journal of the Acoustical Society of America* 127: 1609–1623.

Wyld, Henry Cecil. (1923). *Studies in English Rhymes from Surrey to Pope: A Chapter in the History of English*. John Murray, London.

Zwicky, A. (1976). This rock-and-roll has got to stop: Junior's head is hard as a rock. *Proceedings of Chicago Linguistics Society* 12: 676-697

Kevin McMullin
kevinmcm@alumni.ubc.ca

Using computational models of grammar to assess lexical influences on experimental judgments

Blake Allen
University of British Columbia

This paper presents a novel methodology for addressing a class of questions in phonology: which aspects of phonological knowledge reflect lexical patterns, and which are based on biases independent from the lexicon? The proposed methodology allows direct comparison of grammars constructed based on lexical patterns with grammars constructed based on experimental judgments, as well as the generation of probabilistic predictions about experimental response patterns. As a demonstration of the structure and empirical success of this methodology, this paper considers syllabification judgments in English, corroborating previous studies on the conditioning factors relevant for this phenomenon and—of central importance—determining that only a subset of these factors can originate from grammaticalization of lexical patterns.

1 Introduction

As experimental approaches to phonology have become mainstream, the issue of how to model the relationships between lexical patterns and responses given in experimental tasks has become increasingly important. The most common method of checking for lexically-motivated effects in experimental responses is to run significance tests of some kind on both the lexicon and the experimental data separately (Kawahara 2011; Becker et al. 2012; Moreton and Pater 2012). This approach has proven effective, but only indirectly answers the question at hand, which is whether some lexical pattern has been encoded in the *grammar*. The goal of this paper is to demonstrate that such tests for grammatical sensitivity to lexical trends can also occur at the level of the grammar itself: that using currently existing algorithms, such as the MaxEnt Grammar Tool (Wilson and George 2008), a constraint-based grammar can be constructed from the lexicon of the language in question, and a second grammar can be constructed from experimental results themselves, and that these two grammars can be directly compared with each other. This comparison enables differentiation between lexically-motivated aspects of experimental judgments and aspects with no lexical basis, similar to the traditional approach using multiple significance tests, but appealing to a concrete model of the grammar under investigation.

In this paper I use the topic of lexical patterns' influence on syllabification judgments as an example of the success of this methodology. Eddington et al. (2013) and others have demonstrated that subjects in word division experiments make use of both lexical knowledge, such as which segments constitute well-formed word edges, and "knowledge" or biases with no lexical basis, such as a general preference for V.CV syllabifications over VC.V ones. Several other lexical trends have been proposed as influences on syllabification judgments, but remain untested; these factors are the test cases for the methodology proposed here, with factors like word-edge wellformedness and the preference for V.CV divisions as controls. My comparisons yield a division between lexically-motivated factors and non-lexical factors that compares favorably to existing literature on each of these patterns.

In order to evaluate whether phonological judgments reflect some aspect of lexical knowledge, I begin with a baseline model in which all relevant lexical knowledge is used to make predictions about responses in the experimental task. This model assumes a straightforward relationship among lexical patterns, the phonological grammar, and experimentally-elicited judgments:

(1) *lexicon ⇒ constraint weights ⇒ phonological judgments*

Predicted judgments are generated by constructing a a MaxEnt grammar (Hayes and Wilson 2008)—as implemented in the MaxEnt Grammar Tool (Wilson and George 2008)—from the lexical data, and then applying that grammar itself to the same experimental task as is given to human subjects. If the baseline model is accurate, i.e. if knowledge of all available lexical patterns is used to make phonological judgments, then in an idealized situation we would expect the judgments predicted by the MaxEnt grammar to mirror those resulting from the human experiment. Similarly, since the MaxEnt Grammar Tool enables the creation of a phonological grammar from tableaux representing experimental judgments just as from a lexicon, we would expect in this idealized situation that the grammar (constraint weights) derived from the observed experimental judgments will be equivalent to those "learned" from the lexicon. These relationships can be schematized as in (2).

$$\begin{array}{ccc} \textit{lexicon} & & \\ \Downarrow & & \\ \textit{constraint weights}_{\textit{lexicon}} & \approx & \textit{constraint weights}_{\textit{judgments}} \\ \Downarrow & & \Uparrow \\ \textit{predicted judgments} & \approx & \textit{observed judgments} \end{array} \quad (2)$$

As this paper describes, the judgments and weights on both sides of the diagram above do emerge as generally comparable to each other for the syllabification example case. Testing whether or not a particular factor in the experimental judgments reflects lexical patterns, however, requires evaluating *subset models* of the baseline described above, in which constraints weights are still learned from the lexicon but not all constraints are used to predict judgments in the experimental task. The core logic of this methodology states that if a subset grammar—which excludes all constraint weights pertaining to the factor in question but includes all others—yields predicted experimental results which constitute a better fit to the observed judgments, then the knowledge or bias related to that factor used in the experimental task does not reflect lexical patterns, and so could *not* have been learned from the lexicon. If, on the other hand, removing a particular factor's associated constraints yields a grammar which is able to predict *worse*-fitting responses than those of the baseline grammar, then the relevant phonological knowledge does reflect lexical patterns.

This paper presents the implementational details of this methodology as used to assess the influence of lexical patterns on judgments in syllabification tasks. The following section provides context for the discussion of syllabification judgments and describes factors that previous literature has claimed affect those judgments. Section 3 explains the theoretical underpinnings and technical details of the simulation used to test the origin of knowledge of these factors. Section 4 follows by examining the results of that simulation, and section 5 concludes.

2 The example case: syllabification

Speakers of a natural language possess the ability to judge the relative felicity of different submorphemic divisions, or *syllabifications*, of words in that language, as demonstrated by a variety of experimental tasks (Treiman and Danis 1988; Derwing 1992; Redford and Randall 2005). Native English speakers, for example, prefer to divide the word *alarm* as [ə / lɑɹm] rather than [əl / ɑɹm], whereas judgments about *lemon* are more evenly split among divisions like [lɛ / mən] and [lɛm / ən] (Eddington et al. 2013). These intuitions indicate that the subjects in these experiments use phonological knowledge to assess the acceptability of word divisions, and that some of this knowledge mirrors phonotactic trends observed in the lexicon (Steriade 1999; Hammond 1999; Hall 2006). In order to model the process of learning phonotactic patterns relevant to these word division judgments, however, it is necessary to consider first what phonological structure is responsible for these judgments. The following subsection discusses this issue briefly, and the subsection after that one lays out the predictive factors I will be focusing on in this paper.

2.1 The nature of word division judgments

The core empirical data I consider in this paper are experimentally-elicited judgments about word division. Although such experiments can take many forms, the most typical methodology involves presenting a known word with two alternative pronunciations and asking subjects to choose which pronunciation they prefer; the pronunciation choices have been constructed by adding a brief pause at one possible syllable boundary location (Derwing 1992). See Côté and Kharlamov (2010) for a summary of the apparent effects of different varieties of word division tasks. Although these tasks tend to produce high levels of variability among responses, numerous studies have demonstrated that such judgments reflect principles which can be stated in phonological terms, such as a preference for word divisions which do not produce phonotactically impermissible sequences (Redford and Randall 2005; Eddington et al. 2013).

Due to the high degree of correspondence between such divisions and the boundary locations predicted among the theoretical units called syllables, intra-morpheme word division is typically described as *syllabification* (Itô 1986; Hayes 1989; Becker et al. 2012). Moreover, there is a strong tendency for these word division tasks to create syllable edges that mirror the phonotactics observed at word edges in that particular language. Two types of explanations for this correspondence have been proposed: 1) that word-division judgments make direct use of phonologically-encoded syllable boundaries, whose positions are learned from word-edge phonotactics via Prosodic Licensing (Itô 1986); or 2) that word division produces units that are themselves best considered words, or *sub-words*, and hence subject to the same phonotactic restrictions as "real" words (Steriade 1999; Samuels 2009). Prosodic Licensing, on the other hand, would dictate that because all words must be exhaustively parsed into syllables, any observed word edge will also be a syllable edge.

In this paper, I remain agnostic about which of these interpretations of word division responses is more insightful or accurate. Aside from the possible influence of language-nonspecific constraints on syllable structure, both options yield the same predictions: that the phonotactic knowledge which produces word division judgments will directly reflect the patterns observed at word edges within a particular language. Whether this knowledge is comprised of a set of constraints on allowable syllable structures or on allowable word-edges is immaterial. I describe relevant constraints as simply "edge constraints" rather than syllable- or word-edge constraints, although any of these terms may in fact be appropriate.

Note also that this paper investigates only patterns of word division among words with a single intervocalic consonant, e.g. *lemon* and *lemur* but not *army* or *osprey*. Also, due to the corpus of word division judgments I use, which is data from a binary forced-choice task, there will be no discussion of potential "ambisyllabic" divisions, e.g. [lɛm.mən].

2.2 Posited predictors of word division

Previous research has shown the usefulness of numerous phonological factors in predicting patterns of word division. Because the example simulation described in this paper makes reference to a subset of these, in this section I give a brief description of each factor in turn. I divide them here into control factors, which are factors previously demonstrated to either mirror or ignore lexical trends, and test factors, whose status as lexically-based or not is what my methodology seeks to ascertain.

Steriade (1999) proposed that learning from phonotactic patterns at word edges is the basis for observed patterns of word division, and that the results of this learning can be codified as phonological constraints. Steriade's analysis predicts that word divisions which produce edges that are phonotactically impermissible will be dispreferred, although not rejected outright because of conflicting constraints like ONSET. Dividing the word *lemon* as [lɛ / mən], for example, produces a right edge with an [ɛ], which is never attested in English, and so is less than optimal as a word division. The alternative [lɛm / ən], however, violates both ONSET and NOCODA. Steriade (1999) predicts, then, that neither division will be preferred categorically over the other. This prediction is borne out in the data cited from Derwing (1992) and others, as well as in the experimental results from Eddington et al. (2013). Hence this factor will be taken as the first control in my evaluation of the methodology described here: legality at word edges of particular phonemes in the lexicon of English is known to produce corresponding experimental judgments.

Eddington et al. (2013) describe a large-scale study of word division judgments in English, and their results provide strong evidence that stress patterns can affect word division judgments: a medial consonant tends to associate with whichever of its neighboring vowels is stressed. This finding was also presented by Derwing (1992), among others. They also find that independent of all other factors, there is a tendency toward V.CV divisions over VC.V divisions: as Steriade (1999) described, this indicates a strong influence from the constraints ONSET and/or NOCODA. However, because neither the stress-based effect nor the tendency toward minimizing ONSET/NOCODA violations has ever—to my knowledge—been described as mirroring lexical trends, I take these two factors as the other controls in this test of my methodology. In total there are now three control factors: phonemic word edge legality, which is lexically-derived, and stress-based patterns and the preference for onset consonants, which are not based on lexical patterns.

The following paragraphs introduce the three test factors used here. Two of these three are related to allophonic glottalization. First, Steriade's (1999) theory of syllabification makes use of the observation, cited from Pierrehumbert and Talkin (1992), that word-initial vowels are glottalized. Word divisions which result in unglottalized vowels at the left edge of a "syllable," then, will be dispreferred. Although Steriade does not specify how, in her theory, this vowel glottalization is phonetically implemented, I will follow Dilley et al. (1996) and Redi and Shattuck-Hufnagel (2001) in assuming that glottalized vowels are produced with a constricted glottis during voicing, rather than the insertion of a pre-vocalic glottal stop.

Among English consonants, Hall (2006) claims that only word-final stops are glottalized, as in *cap* [k^{h}æ$^{\text{ʔ}}$p] and *log* [lɑ$^{\text{ʔ}}$g]. In his analysis, avoidance of unglottalized syllable-final stops is responsible for word divisions that associate medial stops with the following vowel. Again, I assume that this glottalization comes in the form of a constricted glottis simultaneous with the stop articulation rather than an inserted glottal stop before or after it.

For the purposes of this paper, I will also make use of the allophonic distinction of anticipatory vowel nasalization, as described by Cohn (1993). The generalization she presents is that vowels immediately preceding nasal consonants in English are anticipatorily nasalized, e.g. *pen* [p^{h}ɛ̃n]. Although Cohn (1993) does not clarify what restrictions might exist on this environment for nasalization, it has been claimed that only coda nasals cause nasalization of their preceding vowel (Ladefoged and Johnson 2010). However, there is no deterministic way to establish the syllabic constituency of English words; this contrast could be considered the cause of Eddington et al.'s (2013) finding that words with a nasal medial (e.g. *lemon*) tend to be divided as VC.V more often than words with an oral stop medial, as that division would prevent a nasalized vowel from occurring at a right-edge (ɛ̃#), as is never found in real English words. To abstract away from this dilemma, I consider all vowels immediately preceding a nasal consonant to be nasalized, and carry out my simulation accordingly.

The list in (3) enumerates the factors introduced in this subsection whose role in word division judgments will be assessed by my simulation, broken down by status as control/test and type of control, if applicable.

(3) **Controls:**

1. phonotactic permissibility of phonemes at word edges (lexical)
2. stress position (non-lexical)
3. ONSET and/or NOCODA (non-lexical)

Tests:

4. word-final stop glottalization
5. anticipatory nasalization
6. word-initial vowel glottalization

3 Learning simulation and model comparisons

Having presented the set of phonological factors being investigated as influences on word division judgments, I now move to the central focus of this paper: describing the methodology I use to simulate phonotactic learning and determine which of these factors can be considered to reflect patterns in the English lexicon. For the purposes of this methodology, I make the following assumptions:

(4)

a. The phonological grammar is comprised of weighted constraints (Legendre et al. 1990).
b. The constraint weights which constitute the phonological grammar are learned from exposure to primary language data (Hayes and Wilson 2008; Hayes 2011a).
c. All subjects in word division judgment tasks have the same phonological grammar, and so variability in responses indicates variable outputs of the grammar, not only inter-speaker variation.
d. If the weight of a constraint that determines word division judgments is learned from phonotactic patterns in the lexicon, then it is learned specifically from the patterns at word edges (see further description below).

The purpose of the assumption in (4d) is only to account in a generalized way for either a) the learning effects of Prosodic Licensing if word divisions are taken to represent syllabifications (i.e. licit syllable edges are learned from word edges) or b) for the claim that if word division can be considered to produce multiple words, then standard word-edge phonotactics will also apply to these. While it is possible that phonotactic constraints not directly related to material at word edges—for example, word minimality effects—could be serving some role in determining word division judgments, I do not take these yet-undescribed factors into account. This restriction is also partly due to the impracticality of running phonotactic simulations of domains larger than word edges. As described later in this paper, even constraints in this restricted category are able to accurately predict word division judgments.

With these assumptions in place, it becomes possible to create a quantitative description (or grammar) in terms of constraint weights both of the English lexicon in general and of a set of attested word division judgments. Under the expectation that word division judgments are driven by the phonological grammar, and that this grammar is itself constructed on the basis of patterns in the lexicon, then in principle these two sets of constraint weights—one based on the lexicon and the other based on word division judgments—should be identical or at least equivalent. Refer to (2) for a schematic diagram of these relationships.

It will come as no surprise, however, that the constraint weights that optimally describe these two sets of data (as calculated using a Maximum Entropy classification algorithm, described below) are not identical, although they are unquestionably similar in many respects. For example, the constraint penalizing [ɪ] at a right edge, *ɪ#, has a high weight in both. Constraint weights will also vary depending on the types of constraints used. For the purposes of this simulation, I encode constraints referring to the presence at a word edge of any phoneme (e.g. *ɛ#), as well as constraints referring to word-edge vowel glottalization (e.g. *ˀV#), stop glottalization (*#ˀC), vowel nasalization (*Ṽ#), stress (*V́#, *V́C#), and also ONSET and NOCODA. Each of these types of constraints is present in both a word-initial and word-final version, and for both polarities (+ and -) of the relevant feature. This includes the constraints ANTI-ONSET and ANTI-NOCODA. A full list of the constraints used in this simulation are provided in section 3.2.2.

Constraint weights are calculated using the MaxEnt Grammar Tool (Wilson and George 2009), an implementation of the maximum entropy classification algorithm. This algorithm has been mathematically demonstrated to converge on the optimal weights to describe any provided set of data (see Hayes and Wilson (2008) for details), and so can be relied upon to produce the correct grammar, or set of weights, given that the attested forms, the available constraints, and each form's violations of those constraints are known. The following subsections describe the specifics of the theoretical underpinnings of phonological grammars as sets of weighted constraints (3.1) and describe the particulars of the implementation of that theory which I use here (3.2).

3.1 Specifics of the theory

In order to make use of maximum entropy classification in determining the optimal constraint weights to describe one of the two target corpora in this study, it is necessary to assume a model of grammar in which knowledge is parametrized as a set of constraints which are weighted relative to each other, and in which grammar outputs can be probabilistic rather than deterministic. This assumption corresponds to the approach taken in Maximum Entropy Grammar (Goldwater and Johnson 2003; Hayes and Wilson 2008; Hayes 2011a). As in Harmonic Grammar (Legendre et al. 1990; Coetzee and Pater 2008) and Optimality Theory (Prince and Smolensky 2008), output forms can either incur constraint violations for particular sequences or forms in an output (Markedness) or for mismatches between elements in the input and elements in the output (Faithfulness). Because all the constraints relevant to word division judgments which I discuss here are Markedness constraints, I do not include Faithfulness constraints in my simulation or this paper.

In weighted constraint models of grammar, the wellformedness of a particular output—that is, its probability of occurring in the language—is the weighted sum of its constraint violations. As an example, consider a language in which nearly all words begin with a consonant and end with a vowel. The grammar that describes (this aspect of) this language may have assign weights of approximately 5 (or any other arbitrary positive number) to the constraints *C# and *#V, and a weight of zero to the constraint *V#. Higher weights indicate that a particular constraint is an accurate descriptor of patterns in the language. Because this language has virtually no words ending with a consonant, the constraint *C#, which incurs violations for forms with word-final consonants and can be thought of as a statement that "this language disprefers word-final consonants," has a high weight. Conversely, because all words in this language end in a vowel, the constraint *V# ("this language disprefers word-final vowels") is highly inaccurate, and thus receives a low weight.

One crucial aspect of harmony scores is that they can be converted into probabilities. To do so, harmony scores are negated and *e* raised to that power, producing a number in the range [0, 1]. This process is performed for every possible form in the language—usually in practice by the sum of these numbers across all *listed* potential outputs—and the [0, 1] value for a given form is divided by that sum. (See Hayes and Wilson (2008) for further discussion.) This method of conversion between harmony scores and probabilities renders it possible to calculate constraint weights *from* the observed probabilities of a set of forms. Given the relative probabilities of each of a set of word shapes, it is possible to calculate the weights of these three constraints which most accurately describe this list of words. This is the process that underlies maximum entropy classification, which is used in the simulation described here.

3.2 Specifics of the implementation

Using the approach outlined above, I calculate the optimal constraint weights for describing a) word-edge phonotactics in the English lexicon and b) experimentally-elicited word division judgments. This subsection lays out the details of my methodology: the corpora I use to represent both data sets, their coding, and the implementation I use of the relevant constraints.

With these corpora and constraints, I use the MaxEnt Grammar Tool created by Wilson and George (2009) to determine the optimal constraint weights for each corpus based on the principle of maximum entropy—essentially "spreading out" the responsibility for attested patterns among as many constraints as possible while still maximizing predictive power, as shown to be phonologically justifiable by Hayes (2011a) and Martin (2011)—and the theory of weighted constraints in (3.1).

3.2.1 Coding of corpora

To represent the lexicon of English that a learner of the language might encounter, I use the Carnegie Mellon University Pronouncing Dictionary (Weide 2005), specifically the corrected subset of entries used in the software BLICK (Hayes 2011a). This lexicon includes 18,033 words from the CMU Pronouncing Dictionary, chosen for having a CELEX (Burnage et al. 1990) frequency greater than zero, not including a productive affix, not being an acronym, etc.

It is crucial for learning accurate English phonotactics that the MaxEnt Grammar Tool is provided not only with the forms in the CMU Pronouncing Dictionary, but also with a list of unobserved—zero-probability—forms. Running the MaxEnt Grammar Tool with only the CMU Pronouncing Dictionary's forms provided yields weights that plainly cannot represent the grammar of English: the constraint *#ŋ, for example, ends up with a *lower* weight than constraints that should have low weights, including *#tʃ and *#θ (2.112, compared to 2.997 and 3.292, respectively). Only when zero-probability [ŋ]-initial forms are provided to the Tool will *#ŋ emerge with an appropriately high weight.

The other difficulty in calculating constraint weights from the CMU Pronouncing Dictionary is its sheer size. 18,033 words is a large number of forms for the MaxEnt Grammar Tool to process, and this would require immense computational capacity and/or time. Adding in all of the (potentially infinite) zero-frequency forms of English would compound the problem further—if it is even possible to represent all possible but unattested pseudo-English words.

For this paper, I use a coding technique which solves both the problem of including non-observed forms and the problem of lexicon size. Because the domain under examination is division of words with only a single intervocalic consonant with no option for ambisyllabicity, and in which only a single vowel flanks this consonant on either side, the only possible responses are *...V.CV...* or *...VC.V....* Accordingly, the only lexical phonotactics which will be relevant to these judgments are those pertaining to the first segment or last segment of a word, in addition to the stress level of the first or last vowel. The compression I performed significantly reduced the number of individual cases for the MaxEnt Grammar Tool to evaluate, and therefore speeds up its run time significantly.

The other advantage to this method of coding the lexicon is that it allows inclusion of the unobserved forms in a straightforward and numerically conservative way. In addition to the sequences which *are* observed in the CMU Pronouncing Dictionary, I added all sequences which are *not* observed there, giving them a frequency of zero, in order to produce the master lexicon to be given to the MaxEnt Grammar Tool, resulting in a manageable 432 entries. This process is parallel to, if more manual than, the inner workings of the UCLA Phonotactic Learner (Hayes and Wilson 2008). The example data in (5) and (6) show the format of original lexicon and the condensed and augmented lexicon, respectively.

(5)

kǽt	1
kɛ́n	1
moʊ́t	1
...	
(length = 18,033)	

(6)

k V́ X X X	2
m V́ X X X	1
X X X V́ t	2
X X X V́ N	1
ŋ V́ X X X	0
X X X V́ h	0
...	
(length = 432)	

This master list of forms was provided to the MaxEnt Grammar Tool as a single "tableau" of forms. Although the different lexical items of English are not in any sense competing outputs of a single input form, providing them this way to the Tool allows it to assess the phonotactic principles underlying the patterns evident in the lexicon. As the following section will explain, this methodology is also empirically validated by the highly accurate weights it produces.

Aside from using the CMU Pronouncing Dictionary to represent the lexicon of English, I adopted the results from Eddington et al.'s (2013) large-scale study of English syllabification as my corpus of attested word division judgments. Parallel to the CMU corpus, I refer to this data set as the BYU corpus, named after the location at which it was elicited. This study collected word division judgments from 842 native speakers of English for a total of 125 words (from a total of five thousand test words) per subject, resulting a grand total of over ten thousand responses. Their methodology was a variant of the "pause-break task" (Derwing 1992), in which subjects seated at a computer screen were visually presented with two possible divisions of each word written in ARPABET and asked to choose one or the other. For example, on the item *lemon*, subjects were asked to indicate a preference for either *L EH / M AH N* or *L EH M / AH N*. Responses were aggregated across all subjects.

For this study, I used only a subset of the words tested by Eddington et al. (2013). Specifically, I removed:

(7)

- words with orthographically double consonant or cluster, e.g. 'kipper', 'badges'
- words in which either side of any syllabification is an independent morpheme, e.g. 'lackey'
- recent or obscure loanwords like 'burro', 'couchant'
- words with multiple common pronunciations, e.g. 'dilute'
- words with more than one intervocalic consonant, e.g. 'escape'

The BYU corpus reflects data from a binary forced-choice task. Accordingly, I provided them to the MaxEnt Grammar Tool as a set of two-output tableaux. Each output form was reduced to only its crucial ...VCV... sub-part, and the the order of two halves of each output word division were reversed so as to allow each one's constraint violations to be assessed using regular expressions. Both the CMU and BYU corpora were provided coded in ARPABET using symbols that correspond to what are sometimes considered the "phonemes" of English, i.e. 24 consonants and 15 vowels, including three diphthongs. Because I am testing for the effects of constraints on allophonic distinctions at word edges, such as glottalization and nasalization, I also coded these lexicons to contain these allophones in the positions suggested by the references listed in section 2.2. For example, all word-initial vowels were marked for glottalization, and all vowels before nasal consonants were marked for anticipatory nasalization.

3.2.2 Coding of constraints

In the absence of any particular theoretical motivation for limiting the number or type of constraints to provide to the weighting algorithm, one approach might be to provide constraints based on all possible featural combinations. The MaxEnt Grammar Tool requires only that all constraints be stated negatively (penalizing forms). Given the input data, the relevant constraints would achieve a high weight and the irrelevant ones would not. However, as with coding the lexicons, there are practical reasons to be prudent in choosing which constraints to allow. Above all is the need to keep the number of constraints to a number that can be assigned weights in a finite and reasonably short period of time.

To this end, rather than use constraints penalizing all possible featural combinations at either edge of a phonological domain, I include two different types of constraints: those on the presence of particular "phonemic" feature sets, and those on the presence of "allophonic" features, both at either the left or right edge[1]. Using both types of constraints is crucial because it allows multiple constraints' violations to

[1] I do not intend to endorse a model of phonology in which there is a strict contrast between phonemes and allophones.

contribute to a form’s (lack of) wellformedness—so-called “gang effects."

The chart in (8) illustrates the phonemic constraints included in the simulation, which are listed as segments but correspond to the appropriate feature values. Allophonic constraints are listed in (9), where only the [+F] versions are listed for economy of space; note that [-F] versions of all constraints are also used in the simulation. Example strings are not all real or even phonotactically permissible words in English.

(8)

constraint name	*description*	*example strings*
*#a$_{\text{phonemic}}$	pick out all allophones of /a/ at a left edge	[a], [ˀaso]
*a$_{\text{phonemic}}$#	pick out all allophones of /a/ at a right edge	[a], [tã]
*#ɔ$_{\text{phonemic}}$	picks out all allophones of /ɔ/ at a left edge	[ɔ], [ˀɔs]
*ɔ$_{\text{phonemic}}$#	picks out all allophones of /ɔ/ at a right edge	[ɔ], [rɔ̃]
*#t$_{\text{phonemic}}$	picks out all allophones of /t/ at a left edge	[t], [ti]
*t$_{\text{phonemic}}$#	picks out all allophones of /t/ at a right edge	[t], [eˀt]
(other “phonemic” constraints elided)		

(9)

constraint name	*description*	*example strings*
*#[+stress]	picks out all left edges with a stressed vowel	[ˀál]
*[+stress]#	picks out all left edges with a stressed vowel	[ní]
*#C[+stress]	picks out all left edges with a consonant and whose first vowel is stressed	[tá]
*[+stress]C#	picks out all right edges with a consonant and whose last vowel is stressed	[ˀáp]
*#[+stopglottalized]	picks out all glottalized stops at a left edge	[ˀpoɪ]
*[+stopglottalized]#	picks out all glottalized stops at a right edge	[teɪˀk]
*#[+vglottalized]	picks out all glottalized vowels at a left edge	[ˀan]
*[+vglottalized]#	picks out all glottalized vowels at a right edge	[sˀeɪ]
*#[+anticnasal]	picks out all nasalized vowels at a left edge	[ũm]
*[+anticnasal]#	picks out all nasalized vowels at a left edge	[sɛ̃]
(and all [-F] versions of these constraints)		

By “phonemic” feature sets, I simply refer to the features of a segment which remain constant across all of its variants I have coded into the CMU and BYU corpora, such as glottalization and nasalization. See Hall (2009) for a discussion of predictability in the relationships among segments.

Following Steriade (1999), I also include ONSET and NOCODA constraints. As their "featural" complements, the constraints ANTI-ONSET and ANTI-NOCODA have been implemented as well.

(10)

constraint name	*description*	*example strings*
ONSET	picks out initial vowels	[ˀa], [ẽɪ]
NOCODA	picks out final consonants	[iˀp], [oʊl]
ANTI-ONSET	picks out initial consonants	[lˀa], [kẽɪ]
ANTI-NOCODA	picks out final vowels	[i], [toʊ]

Using this set of constraints allows the simulation to accurately assess all of the phonological material which can be relevant to word division judgments (according to the assumptions of this paper) while also keeping the number of constraints down to a mere 102: 78 "phonemic" constraints (one for each phoneme, for each edge), 20 "allophonic" constraints (one for each allophonic feature, for each feature value, for each edge), and ONSET, NOCODA, ANTI-ONSET, and ANTI-NOCODA.

The MaxEnt Grammar Tool requires that users provide input files consisting of tableaux with the violations that each form incurs for each constraint already marked. In order to generate these constraint violation matrices, I implemented each constraint as a regular expression, a standard method of string-matching, and used these to automatically produce these matrices.

The set of phonological features used by this simulation was taken from Hayes (2011b). Additions to this feature set were made for each of the above allophones and allophonic features such as glottalization.

4 Results

In this section I give an overview of the results of the simulation described above. 4.1 describes the optimal constraint weights calculated for the experimental word division judgment data set ("the BYU data"), 4.2 provides the weights calculated for the English lexicon ("the CMU data") and describes the changes that need to be made these weights in order to make them comparable to those from the BYU data, and 4.3 gives the CMU-to-BYU fit data that constitute tests for the central question of this paper: which factors that influence word division judgments can be said to reflect patterns in the English lexicon.

4.1 Corroboration of factor relevance

The weights calculated from the responses given in Eddington et al.'s (2013) study at BYU indicate that all of the factors described in 2.2, when expressed as phonological constraints, are active as expected in contributing to those word division judgments, with the possible exception of the stress-based constraints. In the absence of any known method of testing significance levels of the difference between two weights, I compare weights in an absolute fashion, i.e. based only on whether a weight is higher or lower than another. The table in (11) shows a subset of the resulting weights; a full discussion follows.

(11)

*#i	0.5320174414	*i#	0.6891315062
*#æ	0.5768239627	*æ#	1.4975683889
*#[+stopglottal]	0.4387171052	*[+stopglottal]#	0.4387171052
*#[-stopglottal]	0.2324364706	*[-stopglottal]#	0.6449977398
Onset	0.7057193637	NoCoda	0.7168091232
Anti-Onset	0.1717148468	Anti-NoCoda	0.1717148468

Constraints related to legality of particular segments (phonemes) at word edges exhibited constraint weights in line with expectations. For example, constraint weights for segments observed at right edges of words had an average weight of approximately 0.5, whereas constraint weights for segments not observed at right edges were much higher, e.g. *ɛ# had a weight of 1.62. Constraints like *ɛ# had among the highest weights in the set. Note that this does not hold true for a small handful of segments like [h]; the constraint *h# had a weight of only 0.439. This is transparently due, however, to the fact that there is only a single word in the included subset of the BYU data with a medial [h].

The highest weight for the [+/-stopglottalized] series is *[-stopglottalized]# (0.645, as compared to 0.439, 0.232, and 0.438 for the other constraints), as expected on the basis of the lexical generalization that word-final stops are always glottalized. Note that *#[+stopglottalized], which would also be expected to have a high weight, remains low because there are no [+stopglottalized] consonants among the intervocalic consonants in the BYU corpus.

In the [+/-vglottalized] series, the expected bearer of the highest weight, *#[-vglottalized], did emerge with a higher weight (0.706) than other constraints in the series (0.439, 0.301, and 0.310). This result corresponds to the generalization from Pierrehumbert and Talkin (1992) that word-final vowels are glottalized.

The weight of *[+anticnasal]# is higher (0.459) than that of *[-anticnasal]# (0.159), as predicted by the generalization in section 2.2. The *#[+/-anticnasal] constraints, while both of higher weight than *[+anticnasal]#, are not of interest in determining the utility of these constraints, as the nasality of a post-division vowel (e.g. ɔ/gĩ) depends only on the consonant that follows it and is not expected to affect word division judgments.

Unsurprisingly, the constraints ONSET and NOCODA had weights (0.706 and 0.717, respectively) much higher than those of ANTI-ONSET and ANTI-NOCODA (both 0.172).

The stress-based constraints alone are ambiguous in their accordance with their respective empirical finding, i.e. that a V.CV division is more likely when the second vowel is stressed and less likely when the first is stressed. Weights of the right-edge constraints *[-stress]C# and *[+stress]# (0.764 and 0.486) are, as predicted, higher than those of *[-stress]C# and *[+stress]# (0.391 and 0.124). However, the weights across the other two pairs are nearly identical: *#C[+stress] and *#C[-stress] are at 0.303 and 0.307, respectively, and *#[+stress] and *#[-stress] are at 0.574 and 0.570. I interpret these results as indicating that the majority of explanatory power for these stress-based patterns has—for whatever reason—fallen primarily to the right-edge constraints, leaving the left-edge constraints as placeholders whose weights do not greatly affect output choice.

4.2 Standardizing the weight sets

Weights generated from the CMU Pronouncing Dictionary fit rather well to the weights generated from the BYU word division judgment data: the Pearson's product-moment correlation coefficient of the two sets of weights is 0.726 ($p < 1 \times 10^{-132}$). However, the BYU data weights diverge from the CMU data weights primarily in their magnitude, and therefore in their predicted degrees of variability. Whereas weights for the BYU data range from zero to approximately 2.3, weights for the CMU data range from zero to approximately 16. It is not clear whether this is simply because of the high levels of variability in the BYU compared to CMU, or because of the higher token counts in CMU—or both. In any case, these high weights for the CMU data set mean that there is extremely little variation in the word division judgments that they predict: a larger range of weights means that differences between harmony scores will be greater, and hence their differences as log-odds quickly exceed .999 to .001, as exemplified in (12).

(12) *Observed probabilities:*

holy	[...oʊ / li...]	0.81
	[...oʊl / i...]	0.19

Predicted probabilities with raw CMU weights:

holy	[...oʊ / li...]	0.999999574671
	[...oʊl / i...]	$4.2532863268 \times 10^{-07}$

One intuitive solution to this issue would be increasing the effect of the Gaussian prior that forms part of the way each constraint's weight is calculated by the MaxEnt Grammar Tool. Essentially, the lower the sigma value used for the Gaussian prior of a constraint, the more strongly the weight of that constraint will be pulled to some mean value (0 in this case), meaning that a lower sigma will force the range of constraint weights to be smaller. Indeed, decreasing the sigma value for all constraints does decrease the large difference between predicted outputs, but by bringing it low enough ($sigma^2 = 0.001$) to ensure that the ratios of probabilities among output pairs are all within observed ranges, the constraint weights cease to have relative rankings that can be considered representative of English phonotactics. For example, despite the fact that English (and the CMU Pronouncing Dictionary) includes no words that end in [h], drastically reducing the gaussian prior yields a constraint *h# whose weight, 0.15, is *smaller* than that of constraints penalizing frequently observed word-final sequences, such as *ð# (0.261) and *f# (0.214).

My solution is to standardize the CMU constraint weights so that they have the same mean and standard deviation as the BYU weights. Doing so brings the range of predicted variability in word division judgments based on the (standardized) CMU weights to the expected levels and increases the correlation coefficient to 0.814. The predicted probabilities for the two divisions of *holy* (shown in 12) become 0.79 and 0.21, respectively—much closer to the observed probabilities. One alternative, of setting the CMU weights to have the same standard deviation as the BYU weights but a minimum weight of 0, produces negligibly different weights from this solution.

4.3 Factor effects on CMU-BYU correlation

The questions under consideration for this test of the methodology described here are whether each of the allophonic distinctions described in 2.2 correspond to relevant lexical patterns. If the simulations of learning a grammar from lexical patterns and "reverse-learning" a grammar from the experimental response patterns has succeeded, one would expect the control factors to be found to be lexical in origin (word-edge legality) or non-lexical (stress effects and onset maximization). I address these questions by investigating the effect that removing each of these categories of constraints has on the correlation between word divisions predicted from CMU weights (the English lexicon) to those based on the BYU weights (attested word division judgments). The motivating logic of this methodology is that if constraint weights learned from the lexicon do not help predict judgments, then adding those constraints to the set used to predict word divisions will have either no effect or a *negative* effect on the the correlation of those predictions to the observed word divisions, a difference which can be recovered by removing those constraints and checking the change in correlation.

Using the standardized CMU weights, I calculated the Pearson's product-moment correlation coefficient (r) between word divisions predicted by each sets of weights in (13) to those predicted by the optimized BYU weights. Specifically, I correlated the predicted probabilities for the first output of each input form in the BYU corpus; since the sum of the probabilities of each pair of outputs will always be 1, comparing lists containing the predicted probabilities of both outputs for each input form would artificially inflate the correlation percentage. The resulting correlations are as follow in (13). Note that since word-edge legality is plainly an effect of lexical origin, it is excused from this exclusion process.

(13)

excluded constraints	*Pearson's* r
(none)	0.7969
[stress] type	**0.8156**
[stopglottalized] type	0.7428
[vglottalized] type	0.2210
[anticnasal] type	**0.8639**
ONSET, NOCODA, etc.	**0.8088**

Excluding the [stopglottalized] (final stop glottalization) and [vglottalized] (initial vowel glottalization) sets of constraints lowered the correlation between the two sets of predicted word divisions. These results are therefore consistent with the hypothesis that the effect of sensitivity to glottalized vowels and glottalized stops on word division judgments reflects patterns in the English lexicon. Whether this result means that the relevant aspects of the grammar are *learned* from the lexicon is another matter: these points are only evidence that it *possible* that the the grammatical weight of the glottalization contrasts could be learned from exposure to the lexicon of English rather than arising independently from some phonetic or other factor. An alternative explanation—that an external cause is responsible both for responses in the experimental task and for the corresponding patterns within the English lexicon—would predict similar results.

This conclusion contrasts with that for the [anticnasal] (anticipatory nasalization), [stress], and ONSET/NOCODA/etc. sets. Removing each of these sets of constraints increases the correlation between the CMU-based probabilities and the BYU-based probabilities rather than diminishing it, and so their weights appear to introduce unhelpful "noise" into the calculation of word division probabilities. Therefore these biases do not reflect lexical patterns, and may arise from phonetic or other non-phonological factors.

These results are consistent with the predictions made assuming that the methodology introduced here—simulating grammars learning from a lexicon and from experimental data, and comparing those grammars—is a valid one. The stress-based factors and the influence of ONSET and NOCODA show no evidence of reflecting lexical patterns. Moreover, results for the test factors are consistent with existing literature. The finding that influence from anticipatory nasalization is not based on lexical patterns is consistent with Cohn's (1993) conclusion that anticipatory nasalization in English is a purely phonetic, not phonological, process. An additional possibility is that the way I have coded the CMU and BYU lexicons is phonetically inaccurate. The glottalization-related factors, on the other hand, are transparently tied to lexical phonotactics, since word-initial and word-final laryngeal specifications vary cross-linguistically.

5 Conclusion

This paper has presented a simulation of the learning of principles of word division ("syllabification") in English. This simulation was used to exemplify a novel methodology for determining whether patterns observed in experimental data reflect phonological trends in the lexicon. The core of this methodology is the a way to directly compare grammars constructed from a lexicon of forms to grammars calculated from responses in an experimental task. The results of this test simulation indicate that the role of allophonic vowel and stop glottalization patterns constraints correspond to relevant lexical patterns, whereas the role of anticipatory nasalization, stress position, and ONSET/NOCODA must arise from other factors; these findings are consistent with the predictions of existing literature.

As more studies of phonological learning are carried out, methods of assessing whether certain knowledge is a generalization based on observed phonological patterns or an independent bias will be crucial in testing the hypotheses that will emerge. In this paper I have demonstrated that in addition to the usual method of testing for significance of some phonological trend in the lexicon and experimental results separately, comparing the predictions of different sets of weights learned as MaxEnt Grammars can also successfully answer similar research questions.

References

Becker, M., Nevins, A., and Levine, J. (2012). Asymmetries in generalizing alternations to and from initial syllables. *Language*, 88(2):231–268.

Burnage, G., Baayen, R. H., Piepenbrock, R., and van Rijn, H. (1990). *CELEX: A guide for users*.

Coetzee, A. W. and Pater, J. (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language & Linguistic Theory*, 26(2):289–337.

Cohn, A. (1993). Nasalisation in english: phonology or phonetics. *Phonology*, 10(1):43–81.
Côté, M.-H. and Kharlamov, V. (2010). The impact of experimental task on syllabification judgments: a case study of russian. *Handbook of the syllable*, pages 273–94.
Derwing, B. (1992). A 'pause-break' task for eliciting syllable boundary judgments from literate and illiterate speakers: preliminary results for five diverse languages. *Language and Speech*, 35(1-2):219–235.
Dilley, L., Shattuck-Hufnagel, S., and Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24(4):423–444.
Eddington, D., Treiman, R., and Elzinga, D. (2013). The syllabification of American English: Evidence from a large-scale experiment part i. *Journal of Quantitative Linguistics*, 20.
Goldwater, S. and Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Stockholm workshop on variation within Optimality Theory*, pages 111–120.
Hall, K. (2009). *A probabilistic model of phonological relationships from contrast to allophony*. PhD thesis, The Ohio State University.
Hall, T. (2006). English syllabification as the interaction of markedness constraints. *Studia Linguistica*, 60(1):1–33.
Hammond, M. (1999). *The phonology of English: a prosodic optimality-theoretic approach*. Oxford University Press, USA.
Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20:253–306.
Hayes, B. (2011a). Interpreting sonority-projection experiments: the role of phonotactic modeling. In *Proceedings of the 17th International Congress of Phonetic Sciences*.
Hayes, B. (2011b). *Introductory phonology*, volume 36. Wiley-Blackwell.
Hayes, B. and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39(3):379–440.
Itô, J. (1986). *Syllable theory in prosodic phonology*. PhD thesis, University of Massachusetts at Amherst.
Kawahara, S. (2011). Experimental approaches in theoretical phonology. *Rutgers Optimality Archive*.
Ladefoged, P. and Johnson, K. (2010). *A course in phonetics*. Wadsworth Publishing Company.
Legendre, G., Miyata, Y., and Smolensky, P. (1990). *Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations*. Citeseer.
Martin, A. (2011). Grammars leak: Modeling how phonotactic generalizations interact within the grammar. *Language*, 87(4):751–770.
Moreton, E. and Pater, J. (2012). Structure and substance in artificial-phonology learning. *Manuscript, University of North Carolina at Chapel Hill and University of Massachusetts at Amherst*.
Pierrehumbert, J. and Talkin, D. (1992). Lenition of /h/and glottal stop. *Papers in laboratory phonology II: Gesture, segment, prosody*, pages 90–117.
Prince, A. and Smolensky, P. (2008). *Optimality Theory: Constraint interaction in generative grammar*. Wiley-Blackwell.
Redford, M. and Randall, P. (2005). The role of juncture cues and phonological knowledge in English syllabification judgments. *Journal of Phonetics*, 33(1):27–46.
Redi, L. and Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, 29(4):407–429.
Samuels, B. D. (2009). *The Structure of Phonological Theory*. PhD thesis, Cambridge, MA.
Steriade, D. (1999). *Alternatives to syllable-based accounts of consonantal phonotactics*. UCLA.
Treiman, R. and Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language*, 27(1):87–104.
Weide, R. (2005). The Carnegie Mellon pronouncing dictionary [cmudict. 0.6].
Wilson, C. and George, B. (2008). Maxent grammar tool. *Software package. University of California, Los Angeles*.
Wilson, C. and George, B. (2009). Maxent grammar tool (computer software).

Morphophonology and tone in Nata

Andrei Anghelescu
University of British Columbia

In this paper I argue that Nata has an accentual tone system with three tone types which are lexically specified on noun stems. I present evidence that noun stems are the locus of word level tone. I then elaborate an analysis in which the specification of tone type is retained for morphological constructions containing a noun stem. Specifically, the tone type of a noun stem determines the cophonology which will select the output form of a word.

1 Introduction

The primary focus of this paper is the distribution of tone on nominals in Nata. Nata, also known as Kinaata, is a Lacustrine Bantu language spoken in North-western Tanzania (Guthrie's 1971 zone E40). Nouns in Nata fall into one of three tone types. Noun stems are lexically specified for a tone type. In this paper I propose that the tone types determine the accentual pattern of a word; I argue that each type is instantiated by its own cophonology (Anttila 1997). In addition to the analysis of tone types, this paper contributes to the description of nominal morphology in Nata.

In the next section of this paper I describe the structure of nouns. In section 3 I identify three types of words with respect to where high tone is located and argue that tone type is a property of noun stems. In section 4 I lay out an Optimality Theory analysis of the three tone types. The implications of the analysis as well as alternative accounts are discussed in section 5. Finally, in section 6 I summarize the conclusions of this study.

2 Morphology

Nouns in Nata are morphologically complex. They consist minimally of a noun stem affixed with a noun class prefix and an augment; taken together, these three morphemes constitute a word, shown in (1a). The noun class prefix is predictable from the stem and semantics; the augment is generally predictable from the noun class and vowel harmony (Gambarage 2013). These three components are organized as in the schema below (abstracting away from vowel length). I will return to vowel length in noun stems in this section.

(1) Nata nouns

a. Schematic Nata noun

AUG	-PF	-STEM
{V, ∅}	-{CV,∅}	-{CV, CVCV, CVCVCV}

b. Canonical Nata noun

AUG	-PF	-STEM
V	-CV	-{CV, CVCV, CVCVCV}

The majority of noun classes (1/2, 3/4, 7/8, 12/8, 15/6; cf. Johannes (2007)) are cases where the augment is a vowel and the prefix is a CV syllable; for this reason, I refer to this template as the canonical form. The canonical form is shown in (1b). Notable exceptions are class 5 and classes 9/10; these will be discussed in subsection 2.1.

Monosyllabic and disyllabic noun stems, seen in (2) and (3), are both common. Trisyllabic noun stems are attested, but less common[1]. Noun stems may have a long vowel in any position except the final

[1] I will not address quadrisyllabic noun stems. They are attested in the closely related language Ikoma; cf. Aunio

syllable. The inventory of stem shapes is thus:[2]

(2) Monosyllabic stems

a. CV
aa-**ká**
'house'

(3) Disyllabic stems

a. CVCV
ɔ-mɔ́-**rɔrɔ**
'fire'

b. CVVCV
o-mu-**náata**
'Nata person'

(4) Trisyllabic stems

a. CVCVCV
a-**sisíita**
'toothbrush'

b. CVVCVCV
o-mó-**sookani**
'respected person'

c. CVCVVCV
rii-**βurúuŋga**
'egg'

d. CVVCVVCV
a-**taaráari**
'valley'

2.1 Class 5, 9/10 and monosyllables

In this section I will describe the morphology of classes which do not have V-CV syllable structure for the AUGMENT-PREFIX morphemes.

2.1.1 Class 5

Noun class 5 has a CVV prefix, /rii/. This class prefix appears with a phonologically null augment for noun stems longer than one syllable, as shown in (5) . [3] With some monosyllabic stems, the class 5 prefix variably appears with an augment. [4]

(5) C5 prefix with a disyllabic noun stem

∅-	ríí-	βuri
AUG	C5	feather

'feather'

(6) C5 prefix with a monosyllabic noun stem

e-	rii-	só
AUG	C5	eye

'eye'

2.1.2 Class 9

Noun class 9 has a homorganic nasal prefix, /N/, which assimilates to the place of the following obstruent, or otherwise is deleted before non-obstruents. The class 9 augment is variably a short vowel /a/ or a long vowel /aa/. When the noun stem is disyllabic or larger the augment is short, as seen in (7) . When the noun stem is monosyllabic the augment is long, as seen in (8).

(7) C9 prefix with a disyllabic stem

a-	n-	tʃéra
AUG	C9	path

'path'

(8) C9 prefix with a monosyllabic stem

áá-	n-	da
AUG	C9	stomach

'stomach'

(2010).

[2] Noun stems are given in bold. The stems in (2), (3a) and (4a, c, d) are not in canonical noun classes; however, I see no reason to expect that root shapes would be restricted to specific noun classes.

[3] Though not immediately relevant, it is appropriate to make a distinction between morphologically missing augments and phonologically null augments. Instances of the former can be found in Gambarage (2012).

[4] Some, but not all, monosyllabic noun stems surface without an augment in C5. It is not clear what factors contribute to this variability, and it is outside the scope of the current study.

2.1.3 Class 10

Noun class 10 has a CVV augment, /tʃaa/, and a homorganic nasal prefix, /N/, which behaves like the C9 prefix with respect to assimilation and deletion. The structure of Class 10 is exemplified in (9) and (10).

(9) C10 prefix with a disyllabic stem

tʃáá-	n-	tʃera
AUG	C10	path

'paths'

(10) C10 prefix with a monosyllabic stem

tʃáá-	n-	da
AUG	C10	stomach

'stomach'

It is not uncommon in Bantu languages for the class 10 prefix to be used as an additive plural morpheme (Maho 1999). Instead of replacing the class 9 prefix, the class 10 prefix is stacked on top of (to the left of) the class 9 prefix.

However, in Nata the morpheme /tʃaa/ acts like an augment. Specifically, it fails to surface in exactly those contexts in which other augments do not surface (Gambarage 2012). On the other hand, Johannes (2007) points out that in Nata and several related languages, the concord for class 10 is related to the augment, not to the nasal prefix. Therefore, the class 10 augment patterns like both an augment and a(n additive) prefix.[5]

3 Tonology

This section focuses on describing the distribution of high tone on nouns in Nata. I begin by giving some general distributional properties of high tone in Nata. Next, I identify three types of words with respect to where high tone is located. I claim that tone type is controlled by noun stems. The rest of the section examines tone types across morphological contexts, particularly class 5, classes 9/10, and monosyllabic noun stems.

Nata has a two-way contrast in tone. Syllables in Nata are either high-toned, or they are not. High-toned syllables are marked with an acute accent over a vowel (V́). Tone on long vowels is either falling (V́V) or level (V́V́). The canonical Nata word has one high-toned syllable. Each noun belongs to one of three tone types. These types are simply labels which identify the location of high tone in a noun and are marked with a Roman numeral: type I, type II, and type III.

Below I give descriptive definitions for each tone type. Note that these are not the only descriptions which would suffice to capture subsets of the data; however, they are the only descriptions which will capture all of the data.

(11)a. **Tone type I** has a high tone associated to the second syllable of a word unless the first syllable of the word is heavy, in which case the high tone is associated to the heavy word initial syllable.

b. **Tone type II** has a high tone associated to the third syllable of the word.

c. **Tone type III** has a high tone associated to the last syllable of the word.

[5] It may be the case that historically the prefix /tʃa/ was added to an *augmented* class 9 noun.

(i) Historical Class 10 noun

tʃa-	a-	N-	CVCV
C10	AUG	C9	

This would explain why the current augment has a long vowel as well as why it patterns like an augment. Regardless, in modern Nata it is the case that class 10 has the same prefix as class 9, but takes the augment /tʃaa/. Note that the vowel is invariably long for this augment, unlike the class 9 augment.

We can observe all three tone types within a single noun class, as in (12) and (13).

(12) Three types in class 11

a. Type I

o-	rú-	bɛrɛ
AUG	C11	millet

'millet'

b. Type II

o-	ro-	síri
AUG	C11	rope

'rope'

c. Type III

o-	ro-	teeté
AUG	C11	spine

'spine'

(13) Three types in class 7

a. Type I

e-	ɣé-	seku
AUG	C7	door

'door'

b. Type II

e-	ɣe-	síma
AUG	C7	well

'well'

c. Type III

e-	ɣi-	saré
AUG	C7	twin

'twin'

Tone types are not a property of noun class, but rather a property of noun stems. This is demonstrated by the data above: since all three patterns are attested in a single class, there is no evidence that either the class prefix or the augment is selecting for tone type. Therefore, we can generalize that noun class does not correlate with tone type.

We can confirm that tone type is a property of stems by comparing nouns in the singular and plural. In Nata, like many other Bantu languages, noun classes mark (among other things) singular and plural. The plural form of a noun uses a different noun class prefix than the singular form. Additionally, Nata allows the usage of noun class prefixes for semantic effect. Noun class prefixes can be used to create meanings such as diminutive or augmentative (Fortin 2011; Maho 1999). The diminutive is formed by using C5/6 prefixes with noun stems that do not canonically have their singular and plural in those classes.

(14) Type III noun stem across noun classes

a.

e-	ɣi-	saré
AUG	C7	twin

'twin'

b.

e-	βi-	saré
AUG	C8	twin

'twins'

c.

∅-	rii-	saré
AUG	C5	twin

'small twin'

d.

a-	ma-	saré
AUG	C6	twin

'small twins'

The data in (14) demonstrate that a single stem, such as /sare/, retains the same tone type, for this stem type III, across a variety of noun classes.

The data in (15), (16) and (17) demonstrate that all three tone types are consistent across noun classes. Note that type II and type III are neutralized (third syllable = last syllable) in noun class 5 (16c, 17c); I will return to this in the following subsection.

(15) Type I 'millet'	(16) Type II 'plate'	(17) Type III 'twin'
a. o-rú-bɛrɛ c11 (sg)	a. e-ɣe-sóontʃo c7 (sg)	a. e-ɣi-saré c7 (sg)
b. o-βú-bɛrɛ c14 (pl)	b. e-βe-sóontʃo c8 (pl)	b. e-βi-saré c8 (pl)
c. ∅-ríí-bɛrɛ c5 (dim sg)	c. ∅-rii-soontʃó c5 (dim sg)	c. ∅-rii-saré c5 (dim sg)
d. a-má-bɛrɛ c6 (dim pl)	d. a-ma-sóontʃo c6 (dim pl)	d. a-ma-saré c6 (dim pl)

The same three types are observed in stems with three syllables, shown in (18) . We may therefore conclude that tone type does not depend on the number of syllables in a noun stem.

(18) Type I 'waist'	(19) Type II 'old man'	(20) Type III 'deep pan'
a. e-ké-ribitʃi c7 (sg)	a. o-mu-ɣáruka c1 (sg)	a. e-ke-hurerɔ́ c7 (sg)
b. ∅-ríí-ribitʃi c5 (dim sg)	b. ∅-rii-ɣarúka c5 (dim sg)	b. ∅-rii-hurerɔ́ c5 (dim sg)

3.1 Class 5

We have already seen that the noun class 5 prefix, /rii-/, conditions an alternation in tone type I. Class 5 also demonstrates that tone type II must be described as having a high tone on the third syllable of a word, as opposed to the initial syllable of a noun stem. This is best illustrated with trisyllabic noun stems, such as in (21b); in this case, the description of tone type II given in (11b) holds. However, a generalization based on noun stem syllables would fail to capture cases like class 5 where the number of syllables preceding the noun stem has changed.

(21) Class 5 prefix with trisyllabic noun stems

a. Type I			b. Type II			c. Type III		
∅-	ríí-	sooŋgori[7]	∅-	rii-	burúuŋgu	∅-	rii-	reeŋgeetí
AUG	C5	guard	AUG	C5	egg	AUG	C5	blanket
'small guard'			'egg'			'blanket'		

Tone types II and III are indistinguishable with disyllabic noun stems in class 5. Given the descriptions in (11b) and (11c), this is unsurprising. Recall that there is no augment in noun class 5 when the stem is larger than one syllable. Therefore, the third syllable of a word will be the last syllable of the word for noun stems that are disyllabic. The data in (22b) and (22c) illustrate this neutralization.

(22) Class 5 prefix with disyllabic noun stems

a. Type I			b. Type II/III			c. Type II/III		
∅-	ríí-	βuri	∅-	rii-	βaβá	∅-	rii-	muumú
AUG	C5	feather	AUG	C5	wing	AUG	C5	dumbness
'feather'			'wing'			'dumbness'		

[7] cf. o-mó-sooŋgori 'guard'

Since the plural of class 5 is found in class 6, a canonical noun class, we do not find a neutralization of tone type II and III in noun class 6. We can recover the tone type of class 5 nouns in the plural by way of the fact that noun stems are consistently in one tone type. The previously ambiguous tone type II/III noun in (22b) is demonstrated to be in tone type II in (23b).

(23) Class 6 prefix with disyllabic noun stems

a. Type I

a-	má-	βuri
AUG	C6	feather

'feathers'

b. Type II

a-	ma-	βáβa
AUG	C6	wing

'wings'

c. Type III

a-	ma-	muumú
AUG	C6	dumb

'dumbnesses'

3.2 Class 9/10

Because the class 9/10 homorganic nasal prefix is assumed to not be a tone-bearing unit, we can predict that class 9 and 10 will behave like class 5 with tone type II noun stems. Since there is only one syllable preceding the noun stem, tone type II stems will have high tone associated to the second stem syllable, which is the third syllable of the word. This is shown by the data in (24b) and (25b).

(24) Class 9 prefix with trisyllabic noun stems

a. Type I

a-	∅-	súkuβi
AUG	C9	hump

'hump'

b. Type II

a-	ŋ-	gɔkɔ́rɔ
AUG	C9	elbow

'elbow'

c. Type III

a-	∅-	ɲakwaahá[9]
AUG	C9	armpit

'armpit'

We can also observe that tone type I noun stems have high tone associated with the noun stem initial syllable, since it is the second syllable of the word, as in (24a). However, in noun class 10, the same stems have high tone on the first syllable of the word since it is heavy, as in (25a). This demonstrates that we cannot consistently define tone types with reference to morphology.

(25) Class 10 prefix with trisyllabic noun stems

a. Type I

tʃáá-	∅-	sukuβi
AUG	C10	hump

'humps'

b. Type II

tʃaa-	ŋ-	gɔkɔ́rɔ
AUG	C10	elbow

'elbows'

c. Type III

tʃaa-	∅-	ɲakwaahá
AUG	C10	armpit

'armpits'

3.3 Monosyllables

Monosyllabic noun stems display only two tone patterns. For monosyllables in canonical noun classes, tone types II and III will be indistinguishable because the third syllable (type II) is the final syllable of the word (type III). There is no way to tell if a noun stem is of type II or III since we cannot add more syllables to a word.[10]

[9] Example from (Johannes 2007). Some noun stems in class 9, such as a-∅-ɲakwahá and a-∅-ŋgóombe, have an underlyingly prenasalized initial consonant. This is proven by the augmentative forms which are formed via class 14 *e.g.* a-∅-ŋgóombe (C9) ∼ a-ka-ŋgóombe (C14). It could be the case that these were historically vowel initial roots which were reanalysed as consonant initial when used with their noncanonical class prefix.

[10] Locative prefixes are an exception. They can be added on top of a noun class prefix. I will not discuss tone patterns in this context but there are unpredicted complications which I speculate arise due to the interaction between phonology and morphology.

(26) Monosyllabic stems

a. Type I

e-	ké-	ɣi
AUG	C7	wasp

'wasp'

b. Type II/III

e-	ɣi	-kɔ́
AUG	C7	calabash

'calabash'

Below, we see that monosyllabic nouns that take the class 9 prefix maintain the same contrast that other monosyllabic nouns have. Since the augment in class 9 is long with monosyllabic noun stems, tone type I is distinguishable from type II and III. Were the C9 augment to have been a short vowel, as it is with disyllabic noun stems, all three types would be indistinguishable in this context.

(27) Class 9 monosyllabic stems

a. Type I

áá-	n-	da
AUG	C9	stomach

'stomach'

b. Type II/III

aa-	n-	dá
AUG	C9	louse

'louse'

4 Analysis

In this section I will present an Optimality Theory analysis of Nata tone types I, II and III. Specifically, I argue that cophonology theory, as developed by Anttila (1997), captures two crucial properties of nominal tone:

(28)a. Tone type is controlled by noun stem.

b. Tone types take scope over words.

The first property has been discussed in the previous section. I turn now to the second property.

Tone types must be computed not at the level of the noun stem (or prefix, or augment), but rather at the level of the word. This can be demonstrated by the variety of morphemes which can bear high tone in tone type I. Within tone type I, we can contrast canonical noun classes with noun class 9. For canonical noun classes, the second syllable of the word is always the noun class prefix; however, in class 9, the class prefix is not a tone-bearing unit. Therefore, high tone is placed on the first syllable of the noun stem (the second syllable of the word). This illustrates that the scope of tone placement appears to be the whole word as opposed to a particular morpheme.

The chart in (29) schematizes the three realization of the three tone types. For type I, a distinction between heavy and light initial syllables is made. For types II and III, weight distinctions are not made in the chart since they are not relevant for the description of these tone types.

(29) Tone distribution by syllabic structure

Type	Context	σ_1	σ_2	σ_3	σ_4	σ_5	
Type I	(Canonical, C9, 3σ stem)	σ	$\acute{\sigma}$	σ	σ	σ	High tone on the second syllable
	(Canonical, C9, 2σ stem)	σ	$\acute{\sigma}$	σ	σ		
	(Canonical, 1σ stem)	σ	$\acute{\sigma}$	σ			
	(C5,10, 3σ stem)	$\acute{\sigma}_{\mu\mu}$	σ	σ	σ		High tone on the first syllable if heavy
	(C5,10, 2σ stem)	$\acute{\sigma}_{\mu\mu}$	σ	σ			
	(C9, 1σ stem)	$\acute{\sigma}_{\mu\mu}$	σ				
Type II	(All classes, 3σ stem)	σ	σ	$\acute{\sigma}$	σ	σ	High tone on the third syllable
	(All classes, 2σ stem)	σ	σ	$\acute{\sigma}$	σ		
	(All classes, 1σ stem)	σ	σ	$\acute{\sigma}$			
Type III	(All classes, 3σ stem)	σ	σ	σ	σ	$\acute{\sigma}$	High tone on the final syllable
	(All classes, 2σ stem)	σ	σ	σ	$\acute{\sigma}$		
	(All classes, 1σ stem)	σ	σ	$\acute{\sigma}$			

The placement of tone is controlled by the noun stem; however, tone placement has access to whole nouns. This relationship is similar to non-affixal dominance as characterized by Inkelas (1998). Dominant morphemes impose their own cophonology on morphological constructions; recessive forms do not. In Nata all noun stems are dominant and all affixes (noun class prefix, augment) are recessive. I assume that each tone type corresponds to a noun stem being indexed for a particular cophonology. Morphological constructions in which a noun stem is a part are then evaluated by the cophonology for which the stem is specified.

The architecture of cophonologies is simple; there is one divergence from standard OT. Within the master ranking of constraints for a language a specific subset of constraints are re-rankable. This subset is specifically ranked within a cophonology. Under this approach, each tone type will have its own cophonology. The cophonologies are responsible for creating a prominent position corresponding to tone type I, II, or III. I discuss implications and complications of adopting this solution in section 5.2.

Before formalizing the cophonologies which characterize tone types I, II and III, I will argue for an analysis of Nata as a stress system, *i.e.*, a system in which high tone is correlated with the single most prominent position in a word; such an account has previously been proposed by Goldsmith (1987) for other Lacustrine languages with the label "accentual". In this paper, I use the terms "stress system" and "accentual system" interchangeably.

Stress systems have a few well agreed-upon properties (Hyman 2009). Systems that use pitch as stress display at least the following properties:

(30) Definitional properties of stress systems (Hyman 2009)

a. OBLIGATORINESS: every lexical word has at least one syllable marked for the highest degree of metrical prominence
b. CULMINATIVITY: every lexical word has at most one syllable marked for the highest degree of metrical prominence

Nata nouns satisfy both obligatoriness and culminativity. Across all three tone types, every noun has at least one syllable associated with a high tone, but no more than one syllable.[11] In section 5 I will address alternate analyses.

[11] This generalization does not hold in at least one context. In cases where the noun stem is in tone type I and noun class 10, the augment /tʃaa/ is associated with high tone. Such nouns will surface without a high tone in contexts that do not license the augment, such as negation. The syntactic suppression or deletion of the augment in this context must occur after morpho-phonological tone specification.

The first ingredient in an accentual analysis is to ensure that each word has maximally one strong metrical position and that this position is associated with a high tone. The constraints in (31a,b), when ranked above PARSE-σ (31c), will ensure that optimal forms have only one foot. ALLFT-L will force the single metrical foot to be as close as possible to the left edge of a word (tone type I and II). ALLFT-R will force the single metrical foot to be as close as possible to the right edge of a word (tone type III).

(31)a. ALLFT-L: For every foot, the left edge of the foot must coincide with the left edge of a prosodic word. Assign a violation mark for each syllable intervening between the left edge of a foot and the left edge of the prosodic word.

b. ALLFT-R: For every foot, the right edge of the foot must coincide with the right edge of a prosodic word. Assign a violation mark for each syllable intervening between the right edge of a foot and the right edge of a prosodic word.

c. PARSE-σ: Every syllable must be parsed into a foot. Assign a violation mark for each syllable which is not parsed into a foot.

The constraint in (32) associates high tone with metrical heads, where a metrical head is the strong member of a foot, as dictated by the high ranked alignment constraint in (35).

(32) TONE/HEAD: For every metrical head, the left edge of the head must coincide with the left edge of a high tone. Assign a violation mark to any head that is not associated to a high tone.

Because all Nata words have only one prominence, we know that PARSE-σ is ranked below ALLFT-L and ALLFT-R in the master ranking of constraints. TONE/HEAD is never violated and so must dominate DEP-H and *H;[12] it will be omitted in tableaux for the purpose of space. Likewise, PARSE-σ will be omitted from tableaux. The cophonologies for type I and II noun stems will rank ALLFT-L over ALLFT-R; the cophonology for type III will rank ALLFT-R over ALLFT-L.

Type I and type II differ in two ways: type I displays weight-sensitivity (whereas type II does not) and type II displays non-initiality (whereas type I does not). Type I noun stems will have high tone on heavy initial syllable, or else on the second syllable of a word. In contrast, type II noun stems will have high tone on the third syllable of the word regardless of syllable weight. Therefore, I propose that the cophonology for type I noun stems has FOOTBINARITY-μ, (33a), ranked above FOOTBINARITY-σ, (33b). Type II noun stems have FOOTBINARITY-σ ranked above FOOTBINARITY-μ.

(33)a. FOOTBINARITY-μ: A foot consists of exactly two moras. Assign a violation mark to any foot which does not contain exactly two moras.

b. FOOTBINARITY-σ: A foot consists of exactly two syllables. Assign a violation mark to any foot which does not contain exactly two syllables.

Type II noun stems have high tone on the third syllable of a word; we can explain this via the ranking of NON-INITIALITY, as defined in (34), over ALLFT-L in the tone type II cophonology. NON-INITIALITY is undominated for type II stems, but low-ranked for type I and III stems.

(34) NON-INITIALITY: Word initial syllables must not be parsed into feet. Assign a violation mark to any syllable which is leftmost in a word and is parsed into a foot.

[12]DEP-H: Assign a violation mark to any high tone in the output that does not have an input correspondent.
*H: Assign a violation mark to any high tone in the output.

While type I noun stems are weight-sensitive and type II nouns are not, we have evidence that both types prefer iambs (i.e. feet where the strong member is the rightmost). FOOT-TYPE:IAMB is undominated across tone patterns;[13] this constraint will be omitted from tableaux for the purpose of space.

(35) FOOT-TYPE: IAMB: For every foot, the right edge of the foot must be the right edge of a head.

Below I summarize the crucial rankings described above. The constraints in curly braces in (36) represent the subset of constraints which are reranked; the rankings for each cophonology are given in (37). The arrow $x \rightarrow y$ should be read as 'x crucially dominates y'.

(36) Master constraint ranking

FT-TYPE:IAMB, TONE/HEAD,
{ALLFT-L, ALLFT-R, NON-INITIALITY, FOOTBIN-μ, FOOTBIN-σ}
>>
PARSE-σ

(37) Cophonologies for Nata tone types

a. Type I: FOOTBIN-μ ALLFT-L >> ALLFT-R NON-INITIALITY FOOTBIN-σ

b. Type II: FOOTBIN-σ NON-INITIALITY >> ALLFT-L >> ALLRT-R FOOTBIN-μ

c. Type III: FOOTBIN-σ ALLFT-R >> ALLFT-L NON-INITIALITY FOOTBIN-μ

In the type I cophonology, winning candidates never violate FOOTBIN-μ or ALLFT-L. In (38), FOOTBIN-μ and ALLFT-L eliminate the candidates in (38b, c) because they have trimoraic feet which are not aligned to the left edge of the word. We could imagine for an input with only short vowels, such as /o-ru-bɛrɛ/, that candidates could satisfy FOOTBIN-μ but still violate ALLFT-L. An example of such a candidate is [o(rubɛ́)rɛ], which would lose to the optimal [(orú)bɛrɛ].

(38) Type I

o-mu-naata	FTBIN-μ	ALLFT-L	ALLFT-R	NON-INIT	FTBIN-σ
a. ☞ (omú)naata			**	*	
b. o(munáa)ta	*!	*	*		
c. omu(naatá)	*!	**			

For tone type I with in class 5 (or 10), we see that the optimal candidate does not violate either FOOTBIN-μ or ALLFT-L. Any syllabic foot that satisfies left alignment is worse because it would be at least trimoraic and therefore violate FOOTBIN-μ; any moraic foot that violates left alignment is worse than the optimal candidate.

(39) Type I C5

[13] FOOT-TYPE:IAMB is not violated by bimoraic feet where both members have a high tone. I will not address the distinction between tone on short vowels and tone on long vowels, i.e., long vowels with stable high vs. long vowels with falling highs

∅-rii-βuri	FtBin-μ	AllFt-L	AllFt-R	Non-init	FtBin-σ
a. ☞ (ríí)βuri			**	*	*
b. (riiβú)ri	*!		*		
c. rii(βuri)		*!*			

With monosyllabic stems in class 9, the augment surfaces as a long vowel. This long vowel can be parsed into a bimoraic foot as in (39) .

(40) Type I C9 Monosyllabic Stem

aa-∅-ka	FtBin-μ	AllFt-L	AllFt-R	Non-init	FtBin-σ
a. ☞ (áá)ka			*	*	*
b. aa(ká)	*!	*			*
c. (aaká)	*!			*	

In contrast to monosyllabic noun stems, larger stems in class 9 have only one mora preceding the noun stem. Therefore, the first two moras of a word include the first vowel of the noun stem, as in (41a), as opposed to the augment and noun class prefix as in (38a). Morphology aside, the tableau in (41) is identical to the case shown in (38) except for the stem initial long vowel in the latter.

(41) Type I C9[14]

a-N-sukuβi	FtBin-μ	AllFt-L	AllFt-R	Non-init	FtBin-σ
a. ☞ (asú)kuβi			**	*	
b. a(sukú)βi		*!	*		
c. asu(kuβí)		*!*			

Turning now to type II nouns, it is crucial that AllFt-R and AllFt-L are evaluated gradiently. If AllFt-L did not assign a violation mark per syllable intervening between edges, the candidates in (42a,c) would be equal with respect to violations; in that case, AllFt-R would incorrectly select the candidate in (42c) as the winner since it does not violate AllFt-R.

(42) Type II

o-mo-rɛmi	FtBin-σ	Non-Init	AllFt-L	AllFt-R	FtBin-μ
a. ☞ o(morɛ́)mi			*	*	
b. (omó)rɛmi		*!		**	
c. omo(rɛmí)			**!		

In class 9, where the class prefix is not a tone-bearing unit, the augment is not parsed in the optimal candidate due to Non-Initiality. Since FootBin-σ is highly ranked in the type II cophonology, syllables with long vowels are parsed the same as syllables with short vowels. If the type II cophonology ranked FootBin-μ over FootBin-σ, we would expect that the second syllable of the word, /a**taa**raari/, would be parsed into a bimoraic foot in the optimal candidate *[a(taá)raari].

[14]The C9 homorganic nasal does not surface before non-obstruents; see section 2.1.2. I assume that this does not interact with metrical structure.

(43) Type II C9, stem initial long vowel

a-∅-taaraari	FtBin-σ	Non-Init	AllFt-L	AllFt-R	FtBin-μ
a. ☞ a(taaráa)ri			*	*	*
b. (atáa)raari		*!		**	
c. ataa(raarí)			**!		*

The tableau in (44) demonstrates that the type II cophonology selects the type II/III ambiguous form in the correct context; namely for disyllabic noun stems with the class 5 prefix. This follows from the crucial ranking of Non-Initiality over AllFt-L. Since the first syllable of the word is not parsed in the optimal candidate, the last two syllables are parsed into a foot which satisfies AllFt-R, which is highly ranked in the tone type III cophonology. Note that AllFt-R is violated by the winning trisyllabic candidate in (43a).

(44) Type II C5

rii-βaβa	FtBin-σ	Non-Init	AllFt-L	AllFt-R	FtBin-μ
a. ☞ rii-(βaβá)			*		
b. (rii-βá)βa		*!		*	

Compare the tableau in (44) to the tableau in (45); note that both (44a) and (45a) satisfy Non-Initiality and AllFt-R (as well as FtBin-σ). Were the form in (44) put into the cophonology for tone type III, the same optimal candidate would be selected; likewise, were the form in (45) put into the cophonology for tone type II, the same optimal candidate would be selected.

(45) Type III

∅-rii-muumu	FtBin-σ	AllFt-R	AllFt-L	Non-init	FtBin-μ
a. ☞ rii(muumú)			*		*
b. (riimúu)mu		*!		*	

For type III, winning candidates always satisfy FootBin-σ and AllFt-R.

(46) Type III

e-ɣe-teete	FtBin-σ	AllFt-R	AllFt-L	Non-init	FtBin-μ
a. ☞ eɣe(teeté)			**		*
b. (eɣé)teete		*!		*	
c. e(ɣetée)te		*!*	*		*

5 Discussion

In the previous section I identified two key aspects of nominal tone in Nata: first, the tone type of a noun is controlled by the noun stem; second, the tone type of a noun is assigned to the whole word. To capture these two properties I proposed assigning noun stems to a particular cophonology. In this section I will discuss some alternative analyses. I turn first to underlying structural distinctions and then look more carefully at cophonologies.

5.1 Structural approach

The analysis presented in this paper does not use any underlying phonological differences to derive tone types. Instead, being indexed to a particular cophonology will force high tone to be associated to a particular syllable which is metrically determined. In this section I will examine what a structural approach might look like, as well as the difficulties with such an analysis.

A structural analysis requires both a structural distinction, such as an accented syllable, as well as a single mechanism that outputs the observed surface realization given the underlying forms. We might have proposed the following underlying representation for types I, II, and III:

H H

(47) a. Type I: V-CV-[CVCV   b. Type II: V-CV-[CVCV   c. Type III: V-CV-[CVCV

In order to derive the surface forms for tone type I, we could propose a default rule that inserts a high tone onto a toneless word. This rule would need to associate high tone to the first onsetful (CV) syllable or the first syllable with a long vowel ((C)VV).

Types II and III would surface as they are in the canonical noun classes. We would also require a shifting rule that shifts the high tone of type II one syllable to the right in the contexts where there is only one syllable preceding the high tone (C5, 9, 10):

H H

(48) CV.CV → CV.V / #(C)V___

Tone shifting rules have been argued to consist of two distinct steps by Idsardi and Purnell (1997). If the rule were a single step, its iterative application would result in tones "piling up" at the end of a word. A two step spreading rule would require (49a), a rule that spreads a high tone to a toneless vowel to its right, and (49b), a rule that delinks the leftmost vowel from a high tone linked to two vowels.

(49)a. b.

As Idsardi and Purnell (1997) point out, these rules must apply disjunctively by Kiparsky's 1973 Elsewhere Condition. Since the structural description of (49a) is wholly contained within the structural description of (49b), these two rules must apply disjunctively. The rule in (49b) would apply wherever it could, as it is the more specific rule. The rule in (49a) would then apply where it could (but not if the rule in (49b) had applied). This rule ordering would result in surface forms like *[riibúúrúúŋgu] and *[riiβáβá]. This is because the rule in (49a) would never have a chance to apply and delink the relevant tones.

However, to achieve the desired effect of tone shift, the rules must apply such that (50a) feeds (50b). Idsardi and Purnell argue against this analysis of tone shift by way of presenting their own metrical account. Likewise, I will use this observation in favor of the analysis I've presented.[15]

In classic OT, the intermediate representations required by tone shifting rules pose a problem. Cassimjee and Kisseberth's Optimal Domain Theory (ODT) extends feature domains to tone. ODT was developed to account for the same data that had been previously analyzed via metrical tone assignment (Cassimjee and Kisseberth 1992). The current account differs from ODT in that no high tones are assumed to be underlying. High tone is supplied via constraints that build metrical structure as opposed to supplied by the input and then aligned to the edge of a domain. The languages that Cassimjee and Kisseberth (1992) examine allow words with all low toned syllables; Nata does not canonically allow such forms. Under their analysis some extra constraint would be required to guarantee surface forms had a high tone.

5.2 Cophonologies

Cophonologies have been argued to be appropriate accounts of morphologically conditioned phonological variation. The canonical cases include different repair strategies for the same marked sequences across morphemes or morpheme types. This is not the type of case that Nata presents. Instead, I argue that a distinctive property of stems is that they are specified for one of three cophonologies that determine the prosodic shape of words built around those stems.

Nata nouns may contain a stem-stem compound in place of a simplex noun stem. In compound stem contexts, we predict competition between the stem tone types. The simplest prediction is that one member of a stem-stem compound is dominant and consistently determines the tone type.

A second type of conflict that we predict involves non-stem morphemes that assign a tone type. Since stems are proposed to have an inherent tone type, affixes that are specified for some other tone type should interact with the tone type of the stem. Because cophonologies are instantiated by morphological constructions, the outer most affix which is specified for a cophonology should determine the final assignment of high tone. The general prediction is that Nata could have affixes which clearly shift the tone type of a word. This appears to be the case in the verbal system where certain tense and aspect morphology can change the tone type of a verb.

This account predicts that deverbal nouns will inherit the tone behaviour of verb stems they contain. Deverbal nouns use verb stems in place of noun stems; a verb stem includes minimally a verb root and a final vowel. A verb stem may also include extensional suffixes and/or object markers. These constructions offer a testing ground for the predictions made by the present analysis.

6 Conclusion

In this paper I have argued that Nata has an accentual tone system with three tone types which are lexically specified on noun stems. I presented novel data on noun paradigms which illustrate three tone patterns. This data was used to argue that noun stems are the locus of word level tone. I then proposed an analysis in which the specification of tone for a word is determined by the morphological constituents of that word and the tone types for which they are specified. Specifically, the tone type of a noun stem determines the cophonology which will select the output form of a word.

Further work on tone in Nata will extend this model to other morphological constructions. Of particular interest are noun stem compounds, reduplication, deverbal nouns, and verbal extensions. Investigation of other metrical phenomenon, such as minimal words and reduplication, will shed light on the relationship between phonology and morphology in Nata.

[15] If I were to have used different underlying forms for the tone types, tone shifting rules would still fall afoul of the Elsewhere Condition. For instance, if tone type I had a high tone associated to the stem initial syllable, we would require a leftward shifting rule to account for high tone on the prefix in the canonical nouns.

References

Anttila, A. (1997). Deriving variation from grammar. In Hinskens, F., Hout, R. V., and Wetzels, W. L., editors, *Variation, Change and Phonological Theory*, pages 35–68. John Benjamins, Amsterdam.

Aunio, L. (2010). Ikoma nominal tone. *Africana Linguistics*, 16:3–30.

Cassimjee, F. and Kisseberth, C. W. Optimal domains theory and Bantu tonology: a case study from Isixhosa and Shingazidja. ROA 176.

Cassimjee, F. and Kisseberth, C. W. (1992). Metrical structure in Shingazidja. In *Proceedings of the Chicago Linguistic Society*, volume 28, pages 72–93.

Fortin, G. (2011). *The morphology and semantics of expressive affixes*. University of Oxford, Oxford, UK.

Gambarage, J. (2012). Context-of-use of augmented and unaugmented nouns in Nata. UBC qualifying paper.

Gambarage, J. (2013). Tongue root restriction and nominal morphological domains in Nata. UBC qualifying paper.

Goldsmith, J. A. (1987). The rise of rhythmic structure in Bantu. In Dressler, W. U., Luschützky, H. C., Pfeiffer, O. E., and Rennison, J. R., editors, *Phonologica 1984*, London. Cambridge University Press.

Guthrie, M. (1971). *Comparative Bantu*, volume 2. Gregg Press, London.

Hyman, L. M. (2009). How (not) to do phonological typology: the case of pitch-accent. *Language Sciences*, 31:213–238.

Idsardi, W. J. and Purnell, T. C. (1997). Metrical tone and the Elsewhere Condition. *Rivista di Linguistica*, 9:129–156.

Inkelas, S. (1998). The theoretical status of morphologically conditioned phonology: a case study from dominance. In Booij, G. and van Marle, J., editors, *Yearbook of Morphology 1997*, pages 121–155. Springer, Amsterdam.

Johannes, J. G. (2007). The Ki-Nata noun structure. Master's thesis, University of Dar es Salaam.

Kiparsky, P. (1973). "elsewhere" in phonology. In Anderson, S. and Kiparsky, P., editors, *A Festschrift for Morris Halle*, pages 93–106.

Maho, J. F. (1999). *A comparative study of Bantu noun classes*. PhD thesis, Göteborg University.

Andrei Anghelescu
Totem Field Studios 2613 West Mall
Vancouver, BC Canada V6T 1Z4
andrei.anghelescu@alumni.ubc.ca

Hiatus-conditioned accent shift in Japanese:
expanding the vowel sonority hierarchy

Blake Allen
University of British Columbia

Although the majority of verbs in Tokyo Japanese bear accent on their penultimate mora, a small set of verbs exhibit antepenultimate accent, and other verbs freely vary between the two accent positions. In this paper, I investigate these cases of "accent shift" both analytically and experimentally. I propose that such patterns of lexical variability can be accounted for by reference to vowel sonority, using de Lacy's (2004) theory of markedness conflation as the primary analytical tool. I demonstrate that patterns of variability in the Japanese lexicon can be predicted only if a sonority distinction is made between front and back vowels, and only if relevant constraints are able to produce "ganging" effects. This paper also describes a companion experimental study, which strongly suggests that Japanese speakers' knowledge is consistent with the analysis I propose.

1 Introduction

Among the less studied phonological patterns in Standard Tokyo Japanese is a type of accent "shift" conditioned by vowel hiatus in verbs. The lexical items exhibiting this pattern possess an accent (a sharp pitch fall) one mora away from the penultimate position typical of verbal accent. Moreover, many lexical items exhibit phonologically principled variability in their susceptibility to accent shift. In this paper, I propose that this phenomenon—and especially the observed patterns of variability—can be accounted for using a modified version of de Lacy's (2004) theory of *markedness conflation*, in which a stringency hierarchy of vowel sonority levels is encoded as a set of markedness constraints targeting nested subsets of the vowel inventory. After presenting previously undescribed lexical statistics related to accent shift, I argue that the following modifications to de Lacy's theory are warranted: 1) that the sonority scale be expanded to include distinctions between front/back or unrounded/rounded vowels (*contra* Kenstowicz 1997), and 2) that instead of strictly ranked constraints, this hierarchy must be expressed as weighted constraints (Legendre et al. 1990; Hayes and Wilson 2008) in order to allow "gang-up" effects (Farris-Trimble 2008; Jesney and Tessier 2009).

In addition to demonstrating the explanatory adequacy of this proposal, I provide experimental evidence that native speakers of Japanese have knowledge of accent patterns compatible with this theory. To this end, I present the results of an online nonce-word experiment, designed to collect judgments about acceptable accent positions in novel items similar to native Japanese (Yamato) verbs. The results of this experiment indicate that subjects do use a sonority-based generalization compatible with the stringency-theoretic analysis proposed here in their evaluation of acceptable accent locations, and that knowledge of the phonotactic patterns on Sino-Japanese morphemes also plays a large role.

Section 2 introduces the phenomenon of hiatus-triggered accent shift and other relevant phonotactic patterns in the Japanese lexicon. Section 3 then presents the sonority-based analysis of the verbal patterns. Section 4 explains the methodology of a experiment designed to elicit Japanese speakers' generalizations about this pattern, and Section 4 describes its results. Section 5 concludes with a summary and suggestions for future research.

2 Accent in vowel hiatus environments

Hiatus-conditioned accent shift in Japanese verbs is a comparatively under-studied phenomenon. This section presents the basic facts of these patterns, as discussed in Haraguchi (1996), as well as the results of my own examination of accent shift patterns across a comprehensive corpus of native (Yamato) Japanese verbs. I also give a summary of accent patterns in vowel hiatus environments among Sino-Japanese morphemes, which are important in interpreting the experimental results.

2.1 Japanese accent "shift"

In Standard Tokyo Japanese, verbs and verbal ("-i") adjectives are typically considered to fall into two accentual classes: those with accent, which is (almost) always on the penultimate mora or vowel, and those with no accent (McCawley 1965; Haraguchi 1996). Accent is indicated by a sharp drop in F_0 (Beckman and Pierrehumbert 1986). Class membership is not phonologically or semantically predictable, as the overview in (1) sketches. Note that Japanese forms in this paper are provided in phonemic transcriptions, and accent is marked as V́.

(1)

	accented verbs		*unaccented verbs*
bakéru	'to be disguised'	sitoru	'to be soaked'
habámu	'to prevent'	susumu	'to proceed'
simésu	'to show'	tobasu	'to fly (trans.)'

Among the accented verbs and verbal adjectives, accent is typically on the penultimate vowel, as in (1). In a small number of verbs, however, accent falls instead on the antepenultimate vowel:

(2)

	no shift		*shift*
hiéru	'to grow cold'	háiru	'to enter'
koéru	'to become fat'	máiru	'to go (hum.)'
taósu	'to defeat'	káeru	'to return (intrans.)'
kuíru	'to regret'	hirugáesu	'to turn over (trans.)'
	Total: hundreds		*Total: 10*

This phenomenon has been called *accent shift*—parallel to the better known "shift" of accent away from devoiced vowels—by Haraguchi (1996), who to my knowledge is the only other linguist to have documented it. The question of whether a moraically penultimate accent actually shifts leftward during the derivation of these verbs or, instead, there are simply three verbal prosodic classes instead of two will not be a relevant distinction for this paper, and so for purposes of terminological consistency and brevity I will refer to verbs with moraically antepenultimate accent as *accent shift verbs*, and to moraically antepenultimate accent itself as *accent shift*.

All accent shift verbs lack a consonant between the antepenultimate and penultimate vowels, but not all verbs with this -VV(C)V shape exhibit accent shift. Moreover, according to the Japanese Pronunciation and Accent Dictionary (NHK 1985) and Haraguchi (1996), many verbs are variable in their accent shift status. (3) provides a non-exhaustive list of words in this category. All such variably-shifting words have either the vowel sequence [ae] or [oe] in the relevant position; note however that many verbs with these vowel sequences never exhibit accent shift, and almost none exhibit categorical accent shift; exceptions will be discussed presently.

(3)

		variable shift	
osáeru	*or*	osaéru	‘to seize’
otoróeru	*or*	otoroéru	‘to weaken’
kotáeru	*or*	kotaéru	‘to answer’
norikóeru	*or*	norikoéru	‘to climb over’
		Total: 24	

The identity of the penultimate and antepenultimate vowels is the only type of phonological information I have found to correlate with the occurrence of accent shift. Based on the Japanese Pronunciation and Accent Dictionary (NHK 1985) and PSYLEX (Amano and Kondo 1999), I have compiled detailed information about accent patterns on shift-eligible verbs across the all shift-eligible verbs in the native (Yamato) stratum. (5) outlines the observed patterns of shift, non-shift, and variability for each possible vowel-vowel combination, based on the key in (4).

(4) For each antepenultimate-penultimate VV sequence:
- s = there exist verbs with this sequence which exhibit *categorical* accent shift
- v = there exist verbs with this sequence which exhibit *variable* accent shift
- n = there exist verbs with this sequence which *never* exhibit accent shift
- a blank cell indicates that no verbs with this sequence exist

(5)

$V1 \backslash V2$	a	i	u	e	o
a		s		(s)vn	n
i		n		n	n
u		n		n	n
e				n	
o		n		vn	n

As indicated by the *(s)* mark in (5), the number of words with an [ae] sequence which are invariably shifted is limited: only *káeru* ‘to return (intrans.)’, *káesu* ‘to return (trans.)’ and several of their derivative complex verbs, such as *wakagáeru* ‘to be rejuvenated’ and *hirugáeru* ‘to turn over, flutter’. Some derivatives of these words, however, have variable accent shift, e.g. *kutsugáesu/kutsugaésu* ‘to overturn’ and *kurikáesu/kurikaésu* ‘to repeat’[1]. Excepting these special [ae] verbs, then, the chart in (5) can be summarized as demonstrating that: 1) [ai] is always shifted, 2) [ae] and [oe] are either variably or never shifted, and 3) other attested sequences are never shifted.

Non-shifted verbs are more common than variably shifted ones among those with [ae] and [oe] sequences. There are only twenty-four words with variable accent shift: twenty-one with [ae] and three with [oe]. Out of the 15,818 verbs listed in PSYLEX, non-shifted [ae] verbs total to approximately five hundred, and non-shifted [oe] verbs to 108. Words in each category (variably- or non-shifting) range from very high to very low token frequency.

Crucially for the research described here, however, there is a complete lack of information available to Japanese speakers about accent shift on nearly half of the twenty-five possible vowel-vowel combinations; refer back to the empty cells in (5). The origins of these gaps vary. Some, such as the lack of [eV] sequences other than [ee], can be attributed to historical change (Frellesvig 2010). Here I only consider simplex verbs, since verbal compounding can sometimes produce penultimate vowel-vowel sequences, e.g. *deru* ‘to leave’ + *au* ‘to meet’ yields *deau* ‘to happen to meet’. However, accent shift never occurs across

[1] Some derivatives of the ‘return’ words have morphological constituency that is orthographically opaque while others are orthographically transparent, but this distinction correlate with a complex verb’s tendency toward or away from invariable accent shift. By orthographically transparent, I mean that one of the *kanji* (Chinese characters) with the meaning of *kaeru* is present in the compound.

morpheme boundaries, and these compounds often have distinctive accent patterns (Ito and Mester 2007).

Note too that verbs derived from Chinese and assimilated into the Japanese verb system by addition of final *-ru*, *-su*, *-zu*, or *-suru* are also not included here. The subset of these "hybrid" words with adjacent vowels in the penultimate and antepenultimate positions categorically exhibit accent shift, as in *tóoru* 'to go through' and *táisu* 'to oppose'. However, because this shift is actually a universal trait of Sino-Japanese syllables—which never host accent on their second mora—one could argue that these words are uninformative about accent patterns on native Japanese (Yamato) words. As the results of the experiment will demonstrate, however, this distinction cannot be taken for granted: perhaps due to just these such words, Sino-Japanese accent phonotactics cannot be completely disentangled from the accent patterns of words in the Yamato stratum.

2.2 Sino-Japanese accent patterns

Because accent patterns within the Sino-Japanese stratum will prove relevant when discussing the results of my experiment, I provide a brief description of them here. Accent within a Sino-Japanese morpheme will always be on its first mora (McCawley 1965). For the present purposes, the Sino-Japanese morphemes of interest are those with sequences of two vowels in hiatus. The possible shapes of these morphemes derive from the ways that morphemes were borrowed from Chinese languages, as well as the sound changes that occurred to those morphemes; whereas each morpheme in Chinese languages is monosyllabic, their Japanese equivalents are maximally bimoraic. Such sequences found in Japanese are listed below:

	[ai]	*ai* 'love'
	[uu]	*kuu* 'emptiness'
(6)	[ui]	*sui* 'water'
	[ei/ee]	*mei*/*mee* 'name'
	[ou/oo]	*tou*/*too* 'tower'

The sequences [ei/ee] and [ou/oo] are problematic. Whether or not [ei] and [ou] even exist morpheme-internally is a matter of considerable doubt (Kawakami 2001; Hirayama 2003), and at the very least these two pairs of sequences are phonetically identical in such contexts. Members of each pair are only "distinguished" by their different orthographical representations (Okada 1991), which depend on lexical stratum and morphological constituency. The phonetic realizations of these sequences are [ee] and [oo], but whereas they are typically written as [ee] and [oo] in Yamato words, they are written as [ei] and [ou] in Sino-Japanese words, recent loanwords, and a small number of Yamato words. With the exception of loan words, this peculiar orthographical convention is due to historical change (Frellesvig 2010). Even so, because [ei] and [ou] are logically possible vowel combinations in the Japanese phonological system, I include these sequences in my investigation of accent shift.

3 The sonority-based analysis

In this section I present an analysis of Yamato verb accent shift patterns which relies on an extension of de Lacy's (2004) theory of markedness conflation, specifically of vowel sonority levels. Subsection 1 sets up the desired features of such an analysis, and subsection 2 gives the specifics of a constraint-based implementation.

3.1 The generalization

Research on the role of relative vowel sonority tends toward the following sonority scale for a typical five vowel system (Kenstowicz 1997; de Lacy 2004): a > o,e > u,i. The observed vowel sequences in Japanese verbs which exhibit accent shift are, in descending order of accent shift probability, [ai], [ae], and [oe]. Consider the possibility that, just as in the sonority-based stress systems described in Kenstowicz (1997), these patterns can also be described as a tendency for accent to "shift" away from less sonorous vowels and onto more sonorous vowels. The Kenstowicz/de Lacy scale would partition the vowel space as in (7), which is a version of (5) reordered and divided in accordance with this sonority scale.

(7)

$V1\backslash V2$	a	o	e	u	i
a		n	(s)vn		s
o		n	vn		n
e			n		
u		n	n		n
i		n	n		n

The divisions in (7) do not accurately predict the observed "cut-off" points between (variable) accent shift and non-shift: the difference for V_2 between [ae] and [oe] on one hand and [ao] and [oo] on the other exhibits a distinction that is lost by the lack of a partition in this sonority hierarchy between [o] and [e]. Additionally, for V_1, there is a distinction in shift probability between [oe] and [ee]/[ue]/[ie] that is similarly lost. Even if some part of these patterns is attributed to a dispreference for accent shift of sequences of identical vowels (despite a tendency toward just such a shift elsewhere in the language, as described previously), the distinction between [ae] and [ao] at least is evidence for some difference in sonority between [o] and [e].

Despite claims by de Lacy (2006) that neither the roundedness nor the backness of a vowel affects its sonority level, there have been descriptions of phonological processes in at least two languages which indicate phonological sensitivity to just such distinctions. Matthews (1991) gives an analysis of vowel deletion in inter-morphemic vowel hiatus environments in Modern Greek by which the less sonorous vowel deletes. Because of forms like /to + eleɣa/ → [toleɣa] 'I was saying,' he concludes that [o] dominates [e] in terms of sonority. Selection of which vowel deletes cannot merely be positional: compare /ta + evlepa/ → [tavlepa] 'I saw them' and /to + alo/ → [talo] 'the other.' Pulleyblank (2008) adduces from similar patterns in Yoruba evidence for a sonority-based distinction between the high vowels [u] and [i][2]. Phonetic evidence for a difference in sonority between front and back vowels has also been described by Pike (1972) and Foley (1977).

Based on these observations, I propose a more finely articulated hierarchy of vowel sonority than that used by Kenstowicz (1997) or de Lacy (2007):

(8) $a > o > e > u > i$

Assuming such a scale, it is now possible to describe the apparent difference in accent shift probability between [o] and [e] in both V_1 and V_2 positions. Indeed, I propose that there exists some *especially salient* "boundary" in Japanese between [o] and [e] on this scale.

(9) $a > o >>> e > u > i$

The generalization about accent shift patterns based on this revised scale is as follows: sequences crossing the "$o > e$" boundary rightward—i.e., sequences with decreasing sonority in which V_1 is to the left of the boundary and V_2 is to the right—are able to exhibit accent shift. This "conflation" of sonority levels on either side of the $o > e$ boundary parallels the proposal in de Lacy (2007), which will be addressed in the following subsection.

[2]This distinction may also correspond to the difference in markedness between diphthongal *ai* and *au* that Kubozono (2001) describes in Japanese and English.

The most attractive feature of this analysis, however, is its ability to explain the specific patterns of variability among accent shift verbs. Among vowel sequences crossing this boundary which are attested in Japanese verbs, [ai] is always shifted, [ae] is categorically shifted for some words and always shifted for others, and [oe] is only variably shifted. (Some [ae] and [oe] words also never exhibit shift.) These three levels of "shiftability" correspond directly to the distances of each vowel sequence's members from each other on the scale, as illustrated in (10).

(10)

————————————————> [ai] (high shiftability)
————————> [ae] (medium shiftability)
     —————> [oe] (low shiftability)
———> [ao] (non-shifting)
a > o >>> e > u > i

There is exactly one word which fails to exhibit accent shift despite being predicted to do so by this analysis: *oíru* 'to be old'. Note though that this word is of quite low frequency in spoken Japanese, with the word *toshitóru* typically used instead, and that there also exists a medium frequency word *óiru* 'oil' with initial accent but otherwise identical to it. Creating a generalization that does not account for this one form may also be justified by evidence that observation of a single unpredicted form does not drastically change learners' understanding of broader phonological patterns, as in the non-generalization of the *ox* ~ *oxen* paradigm to novel words in English (Becker et al. 2012).

3.2 The constraint-based implementation

This pattern constitutes a case of *markedness conflation* as described by de Lacy (2004). Rather than accent position being sensitive to all distinctions along the (universal) vowel sonority hierarchy as proposed by Prince and Smolensky (2008), the grammar largely ignores certain possible contrasts along that scale, such as [e] > [i], in effect "conflating" these levels. Describing the Japanese patterns is possible using the machinery proposed by de Lacy (2004), and with the addition of an extra layer of distinctions between front and back vowels of the same height ([o] > [e], and [u] > [i]). The subset of this hierarchy of constraints which govern accent placement (rather than the parallel hierarchy of vowels in unaccented moras) is as given below.

(11) *Accent/i >*Accent/i,u > *Accent/i,u,e > *Accent/i,u,e,o

By including all of these constraints, as well as their *Unaccented/V counterparts, it is possible to predict the observed patterns of variability in these sequences. However, it is impossible to do so with a strict ranking of constraints: stringency constraints are unable to produce the necessary "ganging up" effects (Farris-Trimble 2008; Jesney and Tessier 2009). I propose, then, that the phenomenon of accent shift is best described using *weighted* rather than *strictly ranked* constraints, as in MaxEnt Grammars (Hayes and Wilson 2008). Constraint weights must be at approximately the levels indicated in (12); note the use of the > symbol rather than ≫ to indicate that the each constraint in a tier has an appreciably lower weight than any weight of a constraint in the tier above it, rather than strict domination relationships among constraints.

(12)

*V́V
>
Accent/i,u,e , *Unaccented/a,o
>
Accent/i , *Accent/i,u , *Unaccented/a

>
ACCENT/i,u,e,o , *UNACCENTED/a,o,e , *UNACCENTED/a,o,e,u , etc. (inactive)

Because of the way that I have set up these stringency constraints—which, again, are identical to those proposed by de Lacy (2004) with the exception of their reference to backness-based sonority differences—all sequences observed or predicted to exhibit some degree of accent shift will, if given penultimate accent, incur a violation of both *ACCENT/i,u,e and *UNACCENTED/a,o. Although *V́V is given a higher weight than either of these constraints, their combined violation will produce a high enough combined violation weight (a gang-up effect) to allow the antepenultimate accent output to become viable; other sequences will not receive pressure toward antepenultimacy from these constraints, and therefore will always be pronounced with penultimate accent. Among the sequences thereby allowed antepenultimate accent, the degree of acceptability of that accent will depend on the number of accumulated violations of the lower-weighted constraints *ACCENT/i, *ACCENT/i,u, and *UNACCENTED/a: only with all three violated will their combined weight be enough to force categorical accent shift ([ai]), and fewer violations of these produce less antepenultimate accent ([ae] and [oe]). (13) gives example tableaux; note that vertical lines do not indicate strict domination. Winner percentages are given in approximation.

(13)

		*V́V	*ACCENT/i,u,e	*UNACCENTED/a,o	*ACCENT/i	*ACCENT/i,u	*UNACCENTED/a
100%	ái	*					
0%	aí		*	*	*	*	*
70%	áe	*					
30%	aé		*	*			*
40%	óe	*					
60%	oé		*	*			
0%	áo	*		*			
100%	aó			*			*
0%	úi	*	*			*	
100%	uí		*		*	*	
0%	ée	*	*				
100%	eé						
0%	úo	*	*	*		*	
100%	uó						

Although the scope of this paper precludes deeper investigation of the implications of such an analysis, the sketch provided here is sufficiently robust as to make concrete predictions about accent shift in other vowel sequences in Japanese. Specifically, it predicts that [au] will have rather high shiftability, that [ou] will be shifted at approximately the same rate as [ae], that [ui] will never shift, etc. This analysis also predicts that the sequence [oi], despite being found only in one non-shifted word in the Japanese lexicon, will nevertheless exhibit a high degree of shift in novel words. These predictions are summarized in (14), in which a blank cell indicates no shift and a higher number indicates a higher probability of antepenultimate accent.

(14)

$V1 \backslash V2$	a	o	e	u	i
a			2	3	4
o			1	2	3
e					
u					
i					

Because many of these sequences are not found in Japanese verbs, it is impossible to test this prediction by looking at actual Japanese forms. In order to evaluate whether the linguistic knowledge of native Japanese speakers recapitulates the analysis I have described here, it was necessary to carry out a test of accent location judgments on novel Japanese words. The following section describes this test.

4 Methods

Probing phonological judgments with novel ("nonce") stimuli can reveal aspects of linguistic knowledge that would otherwise remain hidden from the researcher (Berko 1958; Zhang and Lai 2008; Becker et al. 2012). The basic protocol of such an experiment—in which native speakers of a language are asked to explicitly or implicitly rate the relative acceptability of different potential forms of novel words—is also appropriate for investigating the question of what generalizations Japanese speakers have formed about accent shift patterns among verbs.

As shown in (5), many of the twenty-five logically possible vowel-vowel sequences are not attested in Japanese verbs, and all of those attested sequences include at least one verb in the accented class. The analysis I have given in the previous section makes strong predictions about the responses native Japanese speakers will offer regarding these novel words which include these unattested sequences, as well as their responses about novel words with observed sequences like [ai] and [ae].

4.1 Procedure

Testing was carried out using an online questionnaire created with Experigen (Becker and Levine 2012). On each testing screen after the instruction frames, subjects first saw a real or novel verb in isolation, as well as two short sentences demonstrating its use in different conjugated forms in order to enhance the sense that these novel words are "real" Japanese verbs, as well as to give information on which conjugation class each verb belongs to. All test novel test items were presented in the *hiragana* orthography. These sentence contexts themselves were also intended to force acceptance of novel forms as Japanese verbs. When finished reading these sentences, subjects pressed a button that played two recorded pronunciations of the target word, which differed only in location of accent: either moraically penultimate or antepenultimate. Subjects were allowed to press the button to listen to the audio tokens as many times as needed. The order of the two recordings was randomized, as was the order of the stimuli overall.

After hearing the target word recordings at least one time, subjects were presented with two buttons with which to indicate their judgments about the relative acceptability of the two pronunciations. These buttons constituted a forced-choice task, in which subjects had to indicate a preference for either the first pronunciation or the second before proceeding. The set of buttons was also labelled as such on each frame.

Subjects additionally provided demographic information and comments at the end of the questionnaire, which were used in conjunction with their judgments about real word accent patterns to determine whether any subjects' responses should be discarded. Twenty-nine subjects participated in total, and three subjects' data were excluded based on failure to accept accent shift on at least one of the lexical items canonically produced with antepenultimate accent (*hairu*, *mairu*, or *kaeru*). These exclusions put the total number of subjects in the result data at twenty-six. No subjects were excluded on the basis of provided demographic information.

Subjects were recruited via postings on student mailing lists, both at the University of British Columbia and at several universities in Japan. Subjects were roughly balanced in terms of gender, and were raised primarily in the central prefectures of the island of Honshuu. Ages ranged from 21 to 47, with a mean age of 34.

4.2 Stimulus design

Stimuli were comprised of recorded productions of 145 words, 125 novel and 20 real Japanese verbs. Recordings were produced by a native speaker of Japanese with training in linguistics. Of the 125 novel words, 25 were distractor items of the form CVCVCV; all other novel words were CVVCV and therefore eligible for accent shift. These 100 words were comprised of four blocks of 25 forms each—one for every possible vowel-vowel combination in Japanese. The four blocks correspond to four different verb conjugations; some conjugations are used more often for integrating loanwords into Japanese, and so it was necessary to control for potential nativeness effects.

Real Japanese words are included in the questionnaire in order to test the reliability of previous literature on their accentual patterns, and as a means of determining which subjects' Japanese may not be close enough to Standard Japanese for their data to be admissible. These words all lacked a consonant between the antepenultimate and penultimate vowels, and included at least one word with each vowel-vowel sequence in this position in Japanese. Some sequences were represented by both an accent shift verb and a standard accent verb, and in the case of [ae], verbs of all three shift-types were included.

Each subject was presented a random set of twelve items out of the twenty-five that comprise each set: the four tested conjugation classes (v-stem verbs, *-nde* verbs, *-tte* verbs, and *-ite* verbs) and distractors. Subjects were also presented twelve random real words out of the set of twenty-two, putting the total number of items per subject at seventy-two. Sessions averaged ten to fifteen minutes in duration.

5 Results

Subjects did not respond in line with the predictions of the sonority-based analysis—at least not in its most literal form. Only 28.8% of subjects indicated a preference for accent shift on novel words with an [ai] vowel sequence, for example, which while significantly greater than such responses for other items, is far from the categorical response that I have predicted. Given this fact, in this section I investigate a *gradient* version of the sonority-based hypothesis in which, despite an overall preference for penultimate accent, sequences with higher predicted shiftability will show a significant trend in that direction compared to other sequences. To look at overall sonority-based patterns irrespective of inter-subject variation, I coded responses as a boolean factor and fit mixed effects logistic regression models, which included random intercepts for each subject, to evaluate the significance of each relevant variable level or combination of levels. A p-value of $< .05$ was taken to indicate significance. All statistical tests were performed using R (R Core Team 2013), and mixed effects models were built using lme4 (Bates et al. 2013).

5.1 Responses on real words

Because the premise of this experiment relies on subjects having knowledge of accent patterns in real Japanese verbs that is comparable to to that found in the Japanese Pronunciation and Accent Dictionary (NHK 1985), it was necessary to test subjects' judgments about real words in addition to those about the novel items. These responses did generally reflect the accent patterns listed in the dictionary: *hairu* and *kaeru* achieved accent shift rates of 100% and 84.6%, respectively, while variable words like *kotaeru* and *otoroeru* had accent shift rates in the 30-45% range. Non-shifted words were, predictably, given zero or near-zero rates of accent shift.

In order to test for significant differences in responses across these three types of words, I constructed mixed effects models of only the real word items, both with the “no shift" response level as the intercept and with the “categorical shift" level as the intercept. These tests indicate that categorically shifted, variably shifted, and non-shifted words do behave differently from each other, as expected, as indicated by $Pr(>|z|)$ values less than .00001.

Subjects do *not* divide neatly into one group that tends to shift variable-accent words and another that does not. Rather, although some subjects tend to prefer shifted accent on more variable-accent words than others, there is no discernible logic behind these responses for each speaker. This finding provides clarification of the reports of “variable accent location” in the Japanese Pronunciation and Accent Dictionary (NHK 1985): rather than some speakers consistently exhibiting shifted accent on some words and other speakers consistently exhibiting default accent on them, each individual speaker understands each word as variable in accent location, although perhaps not uniformly across all variably accented words.

I conclude from the results in this subsection that the assumption underlying my nonce-word study is a valid one: knowledge that native speakers of Japanese possess regarding accent patterns on real Japanese verbs is consistent with the annotations reported by the NHK (1985).

5.2 Responses on novel words

The results on novel word stimuli indicate that participants primarily responded in accordance with the sonority-based hypothesis, and that knowledge of accent patterns in the Sino-Japanese lexical stratum was also called upon during the experiment. Responses indicating a preference for antepenultimate accent are summarized in the table in (15).

(15)

$V1 \backslash V2$	a	o	e	u	i
a	11.6%	6.3%	22.5%	16.9%	28.8%
o	4%	12.5%	6.5%	30.2%	18.2%
e	3.6%	7%	19.1%	7.1%	34%
u	5.8%	9.3%	10.2%	14.6%	16.9%
i	13.7%	3.6%	16.7%	9.6%	7%

One notable aspect of this per-sequence result overview is that the sequence [oi] has been given a relatively high level of accent shift acceptability. This accords with the sonority-based analysis but not with a putative phonologization of *ói based on the word *oíru*. In other words, the general pattern—which I will show to be based on relative vowel sonority—has taken precedence over the effect of this single exceptional item, and so the existence of the single item *oíru* is not sufficient to undermine the analysis proposed here.

In order to test the sonority hypothesis, I fit a logit mixed model with the following predictors:

(16)

1. *sonority difference*: the “raw” sonority difference between V_1 and V_2
2. *crosses o→e*: a binary variable indicating whether or not the sonority levels of the sequence “cross" the crucial boundary I have posited between [o] and [e]
3. *Sino-Japanese seq.*: a binary variable indicating whether or not the sequence is observed in Sino-Japanese morphemes

The sonority difference predictor was set up so as to indicate phonetic difference: it was calculated as (V_1 rank - V_2 rank) on the 5–1 (descending) scale $a > o > e > u > i$, and therefore ranged from 4 ([ai]) to -4 ([ia]). The “crosses o→e” predictor, on the other hand, is included expressly to test the sonority-based analysis I have proposed: it represents the key difference in phonological sonority betwen { a, o } on one hand and { e, u, i } on the other, parallel to the constraint rankings given in (12). Order of presentation (antepenultimate accent heard first or second) and conjugation type were also included, but failed to show strong independent trends in any direction and so are excluded here.

The interaction effect of *sonority difference* and *crosses o→e* corresponds to the predictions of the sonority-based analysis proposed in section 3. Consider the table in (14): the six cells are the vowel sequences for which *crosses o→e* is true. Under a hypothetical analysis in which only this binary variable is responsible for accent shift patterns (comparable to the "pseudo-Japanese" described in the first part of section 3.2), all sequences in one of these six cells would allow accent shift, and the remaining 19 cells would disallow it. Under an account appealing only to raw sonority difference (the *sonority difference* variable), we would arrive at the set of predictions shown in (17). Setting aside the question of whether cells with negative values could be experimentally differentiated from those with zeroes, there are also four cells with positive values which are not present in (14): [ao], [eu], [ei], and [ui].

(17)

V1\V2	a	o	e	u	i
a	0	1	2	3	4
o	-1	0	1	2	3
e	-2	-1	0	1	2
u	-3	-2	-1	0	1
i	-4	-3	-2	-1	0

The interaction effect of *sonority difference* and *crosses o→e* is the subset of (17) that is also in the six cells picked out by *crosses o→e*, and zero for all other cells: exactly the table shown in (**??**). It is possible, then, to determine whether the analysis I have proposed based on stringency constraints is truly a more accurate description of Japanese speakers' knowledge than an account based on raw sonority difference by comparing these two models using R's ANOVA function. This comparison is shown in (18), where the additional degree of freedom in the *stringency-based* model is the interaction effect of *sonority difference* and *crosses o→e*. The effect of being a licit Sino-Japanese intramorphemic vowel sequence, which will be discussed below, is omitted from these models: they are limited to only the sequences that do not occur in Sino-Japanese morphemes. This omission (rather than the addition of an additional factor) is due to the built-in constraints of the lme4 package in R (Bates et al. 2013).

(18)

Model	Degrees of freedom	Log-likelihood	Pr(>Chisq)	
raw sonority difference	8	-253.81		
stringency-based	9	-251.03	0.01831	*

This comparison indicates that the stringency-based predictions are a significantly better fit to the experimental data than a model based on only raw sonority difference, even when the latter model includes a (non-interacting) effect for *crosses o→e*. Limiting the raw sonority difference variable to only non-negative values and comparing these models also yields a significant advantage (Pr = 0.04347) for the stringency-based model.

In order to achieve a fuller picture of which factors are affecting subjects' judgments, it is necessary to look at a single model that includes all of the above variables as well as the (previously excluded) binary *Sino-Japanese seq.* variable. The parameters of the resulting model are shown in (19):

(19)

Coefficient	Estimate	SE	Pr(> $\|z\|$)	
(Intercept: crosses o→e = FALSE, S-J seq. = FALSE)	2.8109	0.45251	5.24e-10	***
sonority difference	0.05293	0.09898	0.59281	
crosses o→e	0.60520	0.76642	0.42974	
Sino-Japanese seq.	-0.88480	0.32026	0.00573	**
*son. difference * crosses o>e*	-0.71602	0.30370	0.01839	*
*son. difference * S-J seq.*	-0.64256	0.24377	0.000839	**
*crosses o→e * S-J seq.*	-1.93931	1.13951	0.08878	
*son. diff. * crosses o→e * S-J seq.*	1.36998	0.45280	0.00248	**

This model is consistent with a combined analysis of vowel hiatus accent patterns which makes reference both to the sonority-based hypothesis described in Section 3 and the accent patterns in Sino-Japanese morphemes. Independent from sonority differences, whether a sequence is found in Sino-Japanese morphemes has a strong correlation with that sequence's accent shift acceptability, suggesting that despite the experiment using only "native Yamato" nonce-words, Sino-Japanese phonotactics were employed in making judgments about them. As the interaction between *sonority difference* and *Sino-Japanese seq.* shows, sonority difference is also predictive of accent shift judgments on sequences that are observed in Sino-Japanese morphemes.

One might wonder why, given the fact that stimuli were presented as Yamato verbs, the phonotactic patterns of Sino-Japanese morphemes proved relevant to subjects' judgments. One likely explanation is that a very limited number of verbs are formed *from* Sino-Japanese morphemes by the addition of a final *-ru*, *-su*, *-zu*, or *-suru*, and accent position on the Sino-Japanese root of these verbs usually remains unchanged. Hence the distinction between Yamato verbs and Sino-Japanese non-verbs is, in practice, of heuristic utility at best. This is especially true considering that many experimental stimuli included vowel sequences like [uu] which are only ever found in Sino-Japanese (or recent loanword) morphemes.

As an additional note, [ei] and [ou] both had a higher accent shift acceptability than even [ai], despite [ai] being categorically shifted both in Yamato words and Sino-Japanese morphemes. I propose that this unexpected result comes about from the influence of orthographical factors: the orthographical sequences <ei> and <ou>, which are phonetically and phonologically equivalent to [ee] and [oo], are found intra-morphemically only in Sino-Japanese words. Seeing items written this way, then, likely caused many subjects to refer to their Sino-Japanese accent pattern judgments, which would dictate that accent is on the first mora of the sequence. Refer back to section 2.2 for details.

6 Conclusion

This paper has presented a novel analysis of hiatus-conditioned accent "shift" in Standard Tokyo Japanese: that the acceptability of antepenultimate accent rather than the more common penultimate accent is directly correlated with the sonority difference between the antepenultimate and penultimate vowels. Specifically, I have claimed that the observed patterns of variability in these accent patterns can best be predicted by de Lacy's (2004) theory of vowel sonority stringency hierarchies and markedness conflation, but only with the crucial additions of phonological sonority distinctions based on vowel backness and the use of weighted constraints rather than strictly ranked ones (Hayes and Wilson 2008). This paper has also described the methodology and results of an experiment designed to test whether native speakers of Japanese use this same principle in their judgments about acceptable accent placement on novel Japanese words, especially those with vowel sequences unobserved in the Japanese lexicon. Analysis of the experimental data indicates that subjects refer to sonority differences in making their judgments, and moreover that this grammatical sonority scale is biased in a way consistent with the analysis proposed here. The experiment also provides evidence that phonotactic knowledge about one stratum (Sino-Japanese) can be put to use even when making judgments about novel forms that appear to belong to another stratum (Yamato).

As a contribution to Japanese phonology in particular, the analysis and experimental data presented here additionally serve to clarify the *loanword accent rule* in Japanese, first proposed by McCawley (1965): "Put an accent on the syllable containing the antepenultimate mora." This rule has since been extended to account for native Yamato accent patterns by Kubozono (2008), but there has been little investigation of which sequences of vowels, especially non-geminate vowels, can be considered tautosyllabic. Assuming that accent position does correlate to syllabic constituency, this paper concludes that there are no categorical criteria for tautosyllabicity among vowel sequences, but rather than syllabicity is evaluated gradiently according to the same phonotactic principles as accent in verbs with vowel hiatus. Perhaps this finding indicates that other phenomena traditionally considered bases for positing phonological syllables in Japanese (Itô 1991; Haraguchi 1996) can also be parsimoniously explained as the result of positional phonotactic restrictions on low-sonority segments.

Appendix: list of stimuli

Stimuli are provided in the Hepburn romanization system.

Nonce-words					**Real words**	
Godan-doushi:			*Ichidan-doushi:*	Distractors:		
-tte	*-nde*	*-ite/-ide*				
taatsu	haabu	kaagu	daaru	nadaku	aogu	'to look up (at)'
yairu	yaimu	kaiku	airu	maniru	aoru	'to fan'
wauru	naubu	saugu	dauru	tabusu	hairu	'to enter'
maetsu	waemu	naeku	yaeru	tasetsu	ieru	'to recover'
saoru	saobu	haogu	waoru	kabomu	kaeru	'to return'
iaru	hiamu	iaku	kiaru	iwaru	kazoeru	'to count'
chiiru	miibu	hiigu	iiru	hiriku	kotaeru	'to answer'
hiuru	kiumu	miugu	hiuru	shimuru	kuiru	'to regret'
nietsu	chiebu	shiegu	chieru	kitemu	mairu	'to go (hum.)'
shiotsu	iomu	iogu	nioru	chigoru	niou	'to smell'
tsuaru	nuabu	muagu	guaru	urasu	oiru	'to age'
nuitsu	kuimu	fuiku	nuiru	tsuniru	shiiru	'to force'
nuuru	tsuubu	suugu	muuru	fuzugu	taeru	'to endure'
tsueru	muemu	nueku	tsueru	nubesu	tooru	'to go through'
fuotsu	suobu	tsuoku	yuoru	tsumoku	ueru	'to starve'
earu	heamu	seagu	hearu	sedaru	ureeru	'to grieve'
neitsu	neibu	heigu	keiru	mebiku	uruosu	'to moisten'
seuru	keumu	meugu	heuru	ebusu	machigaeru	'to be mistaken'
meetsu	neebu	eegu	teeru	nekeru	osaeru	'to restrain'
teotsu	eomu	seogu	georu	hegoru	otoroeru	'to weaken'
soaru	yoabu	toagu	moaru	sobaru	totonoeru	'to arrange'
toitsu	hoimu	noiku	hoiru	moshitsu	uttaeru	'to sue'
koutsu	moumu	hougu	touru	noguru		
noeru	toebu	moeku	toeru	omebu		
nootsu	soomu	toogu	sooru	homoru		

References

Amano, S. and Kondo, T. (1999). *PSYLEX*. Sanseido, Tokyo.
Bates, D., Maechler, M., and Bolker, B. (2013). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999999-2.

Becker, M. and Levine, J. (2012). *Experigen - an online experiment platform*.

Becker, M., Nevins, A., and Levine, J. (2012). Asymmetries in generalizing alternations to and from initial syllables. *Language*, 88(2):231–268.

Beckman, M. and Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology yearbook*, 3(1):5–70.

Berko, J. (1958). *The child's learning of English morphology*. PhD thesis, Radcliffe College.

de Lacy, P. (2004). Markedness conflation in optimality theory. *Phonology*, 21(2):145–199.

de Lacy, P. (2006). *Markedness: Reduction and preservation in phonology*, volume 112. Cambridge University Press.

de Lacy, P. (2007). The interaction of tone, sonority and prosody. In de Lacy, P., editor, *The Cambridge handbook of phonology*. CUP, Cambridge.

Farris-Trimble, A. W. (2008). *Cumulative faithfulness effects in phonology*. PhD thesis, Indiana University.

Foley, J. (1977). *Foundations of theoretical phonology*. Cambridge University Press Cambridge.

Frellesvig, B. (2010). *A history of the Japanese language*. CUP, Cambridge.

Haraguchi, S. (1996). Syllable, mora and accent. In Otake, T. and Cutler, A., editors, *Phonological structure and language processing*, pages 45–75. Mouton de Gruyter., New York, NY.

Hayes, B. and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39(3):379–440.

Hirayama, M. (2003). Contrast in Japanese vowels. *Toronto Working Papers in Linguistics*, 20.

Itô, J. (1991). Prosodic minimality in Japanese. In Deaton, K., Noske, M., and Ziolkowski, M., editors, *CLS 26-II: Papers from the Parasession on the Syllable in Phonetics and Phonology*. Chicago Linguistic Society, Chicago.

Ito, J. and Mester, A. (2007). Prosodic adjunction in Japanese compounds. In *Formal Approaches to Japanese Linguistics*, volume 4, pages 97–111.

Jesney, K. and Tessier, A.-M. (2009). Gradual learning and faithfulness: consequences of ranked vs. weighted constraints. In Schardl, A., Walkow, M., and Abdurrahman, M., editors, *Proceedings of the North East Linguistic Society 38*.

Kawakami, S. (2001). *Nihongo onsei gaisetsu [An outline of Japanese sounds]*. Ofusha.

Kenstowicz, M. (1997). Quality-sensitive stress. *Rivista di Linguistica*, 9:158–188.

Kubozono, H. (2001). On the markedness of diphthongs. *Kobe Papers in Linguistics*, 3:60–73.

Kubozono, H. (2008). Japanese accent. In Miyagawa, S. and Saito, M., editors, *The Oxford Handbook of Japanese Linguistics*, pages 165–189. Oxford University Press, USA.

Legendre, G., Miyata, Y., and Smolensky, P. (1990). *Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations*. Citeseer.

Matthews, P. (1991). *Morphology*. Cambridge: Cambridge University Press.

McCawley, J. (1965). *The accentual system of standard Japanese*. PhD thesis, Massachusetts Institute of Technology.

NHK (1985). Nihongo hatsuon akusento jiten [Japanese pronunciation and accent dictionary]. *Tokyo: Nihon Hoosoo Shuppan Kyookai*.

Okada, H. (1991). Japanese. *Journal of the International Phonetic Association*, 21(02):94–96.

Pike, K. (1972). Phonetics: A critical analysis of phonetic theory and a technic for the practical description of sounds.

Prince, A. and Smolensky, P. (2008). *Optimality Theory: Constraint interaction in generative grammar*. Wiley-Blackwell.

Pulleyblank, D. (2008). Yoruba vowel patterns: asymmetries through phonological competition. In Bobda, A. S., editor, *Explorations into Language Use in Africa*, pages 125–157. Peter Lang, Frankfurt am Main.

R Core Team (2013). *R: A anguage and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Zhang, J. and Lai, Y. (2008). Phonological knowledge beyond the lexicon in Taiwanese double reduplication. *Interfaces in Chinese Phonology: Festschrift in Honor of Matthew Y. Chen on His 70th Birthday. Academia Sinica, Taiwan*, pages 183–222.