HDLS-10

Proceedings of the
High Desert Linguistics Society
Conference

Volume 10

Edited by
Benjamin Anible, Keiko Beers, Laura
Hirrel, Deborah Wager
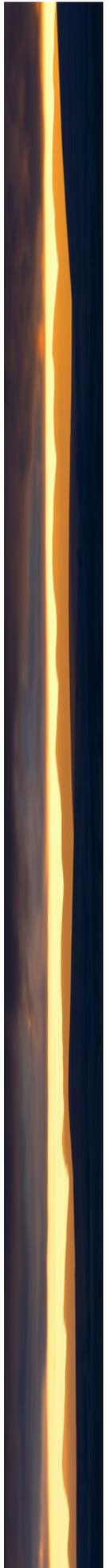
HDLS-10

Proceedings of the
High Desert Linguistics Society Conference

University of New Mexico, November 1-3, 2012

Volume 10

Edited by
Benjamin Anible, Keiko Beers, Laura Hirrel, Deborah Wager

High Desert Linguistics Society
Albuquerque, New Mexico, USA

Acknowledgements

Table of Contents

Voicing and the asymmetry in fricative loan adaptations

Christopher C. Adam and Aaron W. Marks
University of New Mexico

ABSTRACT. In adapting loanword consonants to native phonology, several strategies are available to languages. The sacrifice of place to preserve manner (i.e. manner-faithfulness) and the sacrifice of manner to preserve place (i.e. place-faithfulness) are two such strategies. When languages borrow, they tend to employ one faithfulness strategy more than the other. In languages with an impoverished fricative inventory, it should come as no surprise that fricative manner faithfulness is sacrificed in loanwords in order to accommodate homorganic place faithfulness. While this is a satisfying enough explanation typologically, this study goes a step further by examining the acoustic perceptual reasons for this tendency.*

1. INTRODUCTION. It is commonly observed that when languages borrow vocabulary from one another, the sounds of the borrowed words are mapped in a more or less systematic way to the phoneme inventory of the borrowing language. There is doubtless variation in consistency, given a number of factors such as regularity of borrowing from a given language, the stratal relationship between the languages in question, and spoken contact (as opposed to purely written contact, as with a classical language). However, in cases involving heavy borrowing from a superstrate language, primarily through speech contact, clear patterns arise. Among these are patterns of place-faithful and manner-faithful borrowing.

2. PLACE FAITHFULNESS VERSUS MANNER FAITHFULNESS. To illustrate the contrast between place-faithful and manner-faithful borrowing strategies, it is instructive to observe how the English dental fricatives [θ] and [ð] are adapted to languages employing one or the other strategy. For instance, as seen in Figure 1, words with [θ] are borrowed in Indic languages via substitution of [θ] with the voiceless aspirated dental stop [t̪ʰ], and in Japonic languages via substitution of [θ] with the voiceless alveolar fricative [s]. The English voiced dental fricative [ð] is likewise borrowed in Indic languages via substitution of [ð] with the voiced non-aspirated dental stop [d̪] and in Japonic languages via substitution of [ð] with the voiced alveolar fricative [z].

In terms of faithfulness, Indic languages sacrifice the fricative manner of the English segment in order to preserve the dental place (i.e. place-faithfulness), while Japonic languages sacrifice the dental place of the English segment in order to preserve the fricative manner (i.e. manner-faithfulness).

PLACE FAITHFUL	**UK-US STAND.** [θæŋk] [+ DENT] [+ FRIC] 'thank'	MANNER FAITHFUL
INDIC [tʰæŋk] [+ DENT] [− FRIC]		**JAPONIC** [saŋkɯ] [− DENT] [+ FRIC]

PLACE FAITHFUL	**UK-US STAND.** [ðæt] [+ DENT] [+ FRIC] 'that'	MANNER FAITHFUL
INDIC [dæt] [+ DENT] [− FRIC]		**JAPONIC** [zato] [− DENT] [+ FRIC]
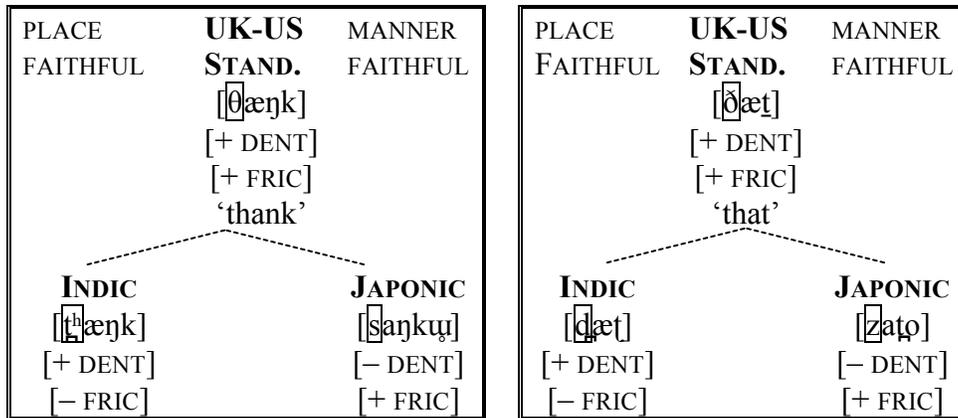
FIGURE 1. Place-faithful and manner-faithful adaptations of interdental fricatives borrowed from English by other languages.

These patterns of place and manner faithfulness are typical not only of synchronic patterns of word borrowing, but also of diachronic patterns of vocabulary inheritance. For instance, as seen in Figure 2 below, the Proto-Semitic voiceless dental fricative *[θ] continues into Arabic with faithful adherence to both place and manner, while it evolves into the Aramaic voiceless dental stop [t̪] with faithful adherence to place but not manner, and it evolves into the Canaanite voiceless post-alveolar fricative [ʃ] with faithful adherence to manner but not place. Likewise, the Proto-Semitic voiced dental fricative *[ð] continues into Arabic with faithful adherence to both place and manner, while it evolves into the Aramaic voiceless dental stop [d̪] with faithful adherence to place but not manner, and it evolves into the Canaanite voiceless post-alveolar fricative [z] with faithful adherence to manner but not place.

PLACE FAITHFUL	**PROTO-SEMITIC** *θ-l-θ [+ DENT] [+ FRIC] 'three'	MANNER FAITHFUL
ARAM [t̪la:t̪a:] [+ DENT] [− FRIC]	**ARABIC** [θælæ:θæ] [+ DENT] [+ FRIC]	**CANAAN** [ʃlo:ʃa:] [− DENT] [+ FRIC]

PLACE FAITHFUL	**PROTO-SEMITIC** *ð-b-ḥ [+ DENT] [+ FRIC] 'sacrifice'	MANNER FAITHFUL
ARAM [d̪əbaḥ] [+ DENT] [− FRIC]	**ARABIC** [ðabaḥ] [+ DENT] [+ FRIC]	**CANAAN** [za:baḥ] [− DENT] [+ FRIC]
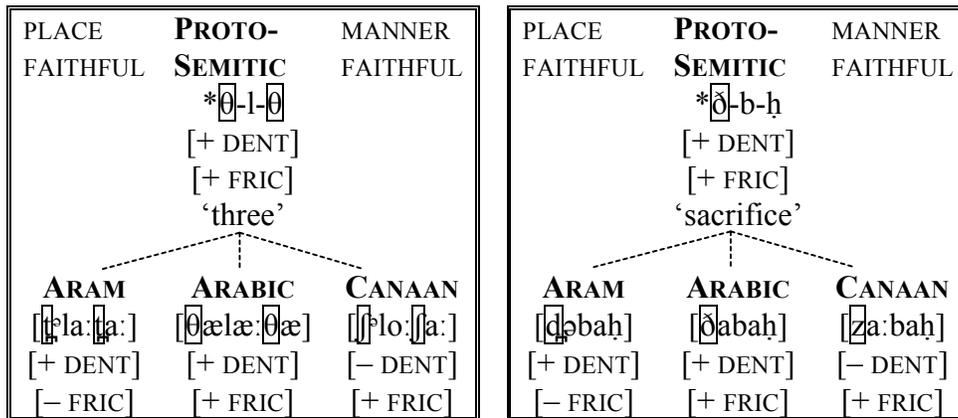
FIGURE 2. Place-faithful and manner-faithful adaptations of interdental fricatives inherited from Proto-Semitic by descendant Semitic languages.

In terms of faithfulness, Aramaic sacrifices the manner of the Proto-Semitic segments to preserve their place, while Canaanite sacrifices the place of the Proto-Semitic segments to preserve their manner. Arabic remains faithful to both the place and manner of the original Proto-Semitic segments.

The history of Arabic dialects replicates these results. For instance, as seen in Figure 3 below, the Classical Arabic voiceless dental fricative [θ] emerged in Syrian Arabic as the voiceless dental stop [t̪], and was borrowed into Persian as the voiceless alveolar fricative [s]. The Classical Arabic voiced dental fricative [ð] emerged in Syrian Arabic as the voiced dental stop [d̪], and was borrowed into Persian as the voiced alveolar fricative [z].
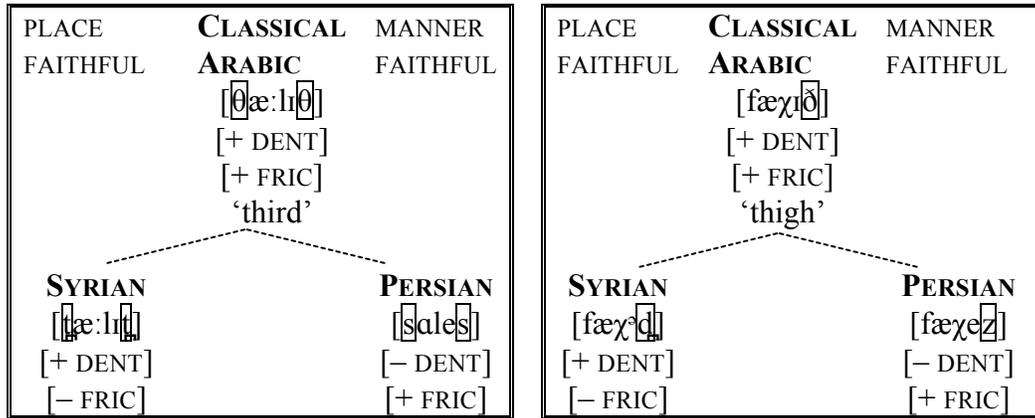
FIGURE 3. Place-faithful and manner-faithful adaptations of interdental fricatives inherited from Classical Arabic by languages with derivative vocabularies.

In terms of faithfulness, Syrian Arabic sacrifices the manner of the Classical Arabic segment to preserve its place, while Persian sacrifices the place of the Classical Arabic segment to preserve its manner.

These results are replicated in Late Koinê Greek, which also had the interdental fricatives. For instance, as seen in Figure 4, the Greek voiceless dental fricative [θ] was borrowed into Georgian as the voiceless aspirated dental stop [t̪ʰ], and into Russian as the voiceless labiodental fricative [f]. The same changes are found in the English voiceless dental fricative [θ] emerging as [f] in several UK and US dialects of English—for contrastive purposes, the Indic English realization of [θ] as [t̪ʰ] is repeated.
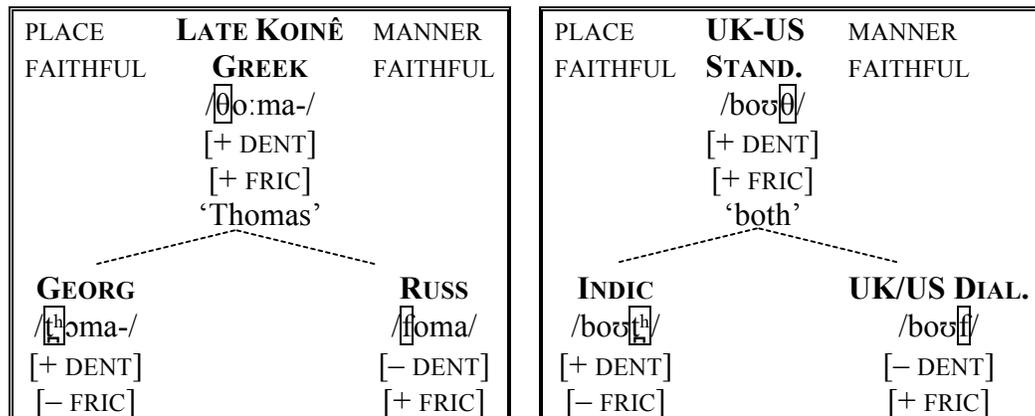
FIGURE 4. Place-faithful and manner-faithful adaptations of interdental fricatives borrowed from Greek by Eastern-Christendom Heritage languages and inherited from Standard English (US/UK) by dialects with derivative vocabularies.

In terms of faithfulness, Georgian (like Indic) sacrifices manner to preserve place, while Russian and certain US/UK English dialects sacrifice place to preserve manner.

3. EARLIER LITERATURE DEALING WITH THE QUESTIONS WE ADDRESS. Relatively little literature has used a faithfulness-based approach to address the varied adaptations of borrowed interdental fricatives across languages. Hock (1991) and Kim (2009) claim that borrowing languages tend to find a one-to-one mapping of foreign sounds onto the native phoneme inventory. Hock adds that the mapping of non-native fricatives onto aspirate stops occurs 'because the friction noise of these aspirates approximates the acoustic impression of the foreign fricatives' (Hock 1991:394).

What Hock (1991) does not address is that it is not merely the similarity of the target fricative's spectral characteristics and those of the aspirate burst, but also the mean duration of the segment which determines mapping, as the latter explains why voiced fricatives are not borrowed as voiced aspirates but rather as voiced non-aspirates.

Hock comments that 'system-based nativizations like these may be quite rare' (394), evidently not recognizing the current study's findings that such systematic nativizations do, in fact, occur regularly on any given side of a typological divide, which is itself determined by a given language's inventory.

In Adler's 2006:1027 study of English fricatives borrowed into Hawaiian, she notes a manner-faithful tendency, pointing out that the 'English coronal plosives are frequently adapted through a change in place of articulation' rather than a change in manner, and similarly that English fricatives are frequently adapted through a change in place of articulation rather than a change in manner (1028). Accordingly, non-native coronal plosives tend to velarize, as with the English alveolar [t] becoming the Hawaiian velar [k] in [kiníkə] 'tinker', while non-native fricatives tend to converge on the language's sole fricative [h], as with the English alveolar [s] realized as the Hawaiian glottal [h] in [kəlíhi] 'crease' (Adler 2006:1028), both replicating identical shifts in the Proto-Polynesian transition to Hawaiian (Elbert 1964:400), though Adler doesn't discuss this parallel.

Adler (2006:1028) claims that borrowed interdental fricatives produced surprisingly inconsistent outputs in Hawaiian, with some borrowed interdental fricatives realized as either a glottal fricative, as in the case of other borrowed fricatives, or as a velar plosive as in the case of coronal plosives. This should come as little surprise, however, since in the spoken form of English used by bilingual Hawaiian speakers, the interdental fricatives [θ] and [ð] are pronounced as plosives [t] and [d], respectively (Sakoda & Siegel 2004:743).

Cowell (1964:252) also discusses varied treatment of intervocalic fricatives adapted to dialects of Arabic which lack them. In such dialects as Syrian, they are pronounced place-faithfully in speech-register words, as with the [ð] of ʔaxað 'take' becoming [ɖ] in ʔaxaɖ, and they are pronounced manner-faithfully in literary-register words, as the [ð] of ittaxað 'undertake' (a derived form of the foregoing) becoming [z] in ittaxaz.

Gonet and Pietroń (2006) discuss a different reason for such apparent inconsistencies. In the case of Polish, speakers tend to use place-faithful strategies for the interdental fricative [ð] in pre-vocalic positions but manner-faithful strategies in pre-consonantal positions.

Dinssen and McGarrity (1999:175–177) discuss distributionally based diversity of realization in English-speaking children with speech impediments, such as dental lisps, and they theorize how derived forms of words affect variation between [s] and [θ] in words with underlying [s]. Dinssen and Gierut (2009:442) note cases of speech impediments which collapse [s] and [θ] place-faithfully in production, though they retain a distinction of them in perception.

Paradis and LaCharité (2012) do not discuss place-faithful and manner-faithful borrowing in typological terms, but instead discuss the diversity in evaluative terms, claiming that place-faithful borrowings of interdental fricatives [θ] and [ð] as [t] and [d] represent a cross-linguistically proper 'phonological adaptation' (99), while manner-faithful borrowings of interdental fricatives [θ] and [ð] as either [s] and [z] or as [f] and [v] are based on 'faulty perception' (99) or a 'flawed production-based attempt' (100) because speakers find foreign interdental fricatives to be 'difficult to produce' (108).

Park (2007:18–19) approaches the question from an OT perspective but relies on an independently motivated generation of a [*CONT] constraint which disprefers continuants in place-faithful languages. The major problem with this constraint is that claiming a language has a constraint avoiding the most common fricative in the world's languages, [s] (Maddieson 1984:44), ignores the question of how a given place-faithful language would have any [s] distribution at all, if such a constraint exists.

Ahn (2003) presents a distribution-based argument which poses problems mostly with the data under analysis. The most notable problem in this work is the claim that Japanese is place-faithful (like Ahn's native Korean), even though common borrowings show Japanese to be exemplarily manner faithful,[1] and even though other linguists working from the perspective of explaining Korean agree that Japanese is manner-faithful (Park 2007:6).

4. PHONEME INVENTORIES. For the present study, we examined a range of stop and fricative phonemes, including voiced/voiceless oppositions, and (for the stops), aspirated/non-aspirated oppositions, across five places of articulation: labial, dental, post-dental, palato-alveolar, and dorsal. Table 1 gives the full range of segments analyzed.

CONSONANTS	LABIAL		DENTAL		POST-DENT		PALAT-ALV		DORSAL	
STOP	p	b	t̪	d̪	t	d	tʃ	dʒ	k	g
ASP STOP	pʰ	bʱ	t̪ʰ	d̪ʱ	tʰ	dʱ	tʃʰ	dʒʱ	kʰ	gʱ
FRICATIVE	f	v	θ	ð	s	z	ʃ	ʒ	x	ɣ

TABLE 1. Segments under Analysis.

Table 2 gives the phoneme inventory of the Tibeto-Burman language Newari (Genetti 2007). Newari is rich in stop contrasts, but has only two phonemic fricatives, /s/ and /ʃ/.

CONSONANTS	LABIAL		DENTAL		POST-DENT		PALAT-ALV		DORSAL	
STOP	p	b	t̪	d̪	t	d	tʃ	dʒ	k	g
ASP STOP	pʰ	bʱ	t̪ʰ	d̪ʱ	tʰ	dʱ	tʃʰ	dʒʱ	kʰ	gʱ
FRICATIVE					s		ʃ			

TABLE 2. Phoneme Inventory of Newari.

Table 3 gives the phoneme inventory of the Indo-European language Persian (Payne 2009). Persian is relatively rich in fricatives, but has no minimal contrasts of homorganic stops, save a voicing contrast.

Consonants	Labial		Dental		Post-Dent		Palat-Alv		Dorsal	
Stop	p	b	t̪	d̪			t͡ʃ	d͡ʒ	k	g
Asp Stop										
Fricative	f	v			s	z	ʃ	ʒ	x	ɣ

TABLE 3. Phoneme Inventory of Persian.[2]

As seen in Table 4, Newari has a poor inventory of fricatives and must adapt foreign fricatives to conform within the limits of its stop series. In contrast, as seen in Table 5, Persian has a rich inventory of fricatives and must adapt foreign fricatives to conform within the limits of own fricative series.

[f] > [pʰ]	Arabic [fæwd͡ʒ]	>	Newari [pʰod͡ʒ]	'army'
[x] > [kʰ]	Arabic [xæzæ:næ]	>	Newari [kʰədʒana]	'treasury'
[v] > [b]	Perso-Arabic [væqt]	>	Newari [bəkʰət]	'time'
[z] > [d͡ʒ]	Arabic [xæzæ:næ]	>	Newari [kʰəd͡ʒana]	'treasury'

TABLE 4. Mapping of non-native fricatives borrowed into Newari.

[θ] > [s]	Arabic [θæ:lɪθ]	>	Persian [sɑles]	'third'
[ð] > [z]	Arabic [ʔɪðæ:n]	>	Persian [ezɑn]	'call to prayer'

TABLE 5. Mapping of non-native fricatives borrowed into Persian.

5. ASYMMETRY IN BORROWED VOICED AND VOICELESS FRICATIVES. There appears to be an asymmetry in the mapping of fricatives to languages possessing voiced/voiceless and aspirated/non-aspirated contrasts in their stop series: voiceless fricatives are mapped onto voiceless aspirated stops, while voiced fricatives are mapped onto voiced non-aspirated stops. This asymmetry is seen across a wide range of source languages and time periods, and is by no means negligible. Table 6 gives some examples in Hindi.

	Imported Fricative	Maps to Hindi Segment	Source	Hindi pronunciation	Meaning
Voiceless	x	kʰ	[xun] (Persian)	[kʰun]	'blood'
	θ	tʰ	[miθeɪn] (UK English)[3]	[mɪtʰen]	'methane'
	f	pʰ	[fæɪṣæl] (Perso-Arabic)	[pʰɛsla] / [fɛsla]	'decision'
Voiced	ɣ	g	[ɣælæt̪] (Perso-Arabic)	[gələt̪lə 'wrong'	
	ð	d	[hæðəɹ gɹæm] (English)[4]	[hedər grahəm]	'Heather Graham'
	v	b	[vən] (Sanskrit)	[bən]	'forest'

TABLE 6. Mapping asymmetry among imported fricatives in Hindi.

6

We sought an explanatory factor for this asymmetry, and turned to quantifiable acoustic measurements. We tested the hypothesis that the asymmetry in the segment mapping occurs on the basis of *mean duration*.

For example, we hypothesized that the mean duration of the imported voiceless velar fricative [x] is closer to that of the native voiceless aspirated velar stop [kʰ] than it is to the native voiceless non-aspirated velar stop [k]. On the other hand, we hypothesized that the mean duration of the imported voiced velar fricative [ɣ] is closer to that of the native voiced non-aspirated velar stop [g] than it is to the native voiced aspirated velar stop [gʱ].

In order to confirm this hypothesis, we apply a chi-squared statistical significance test to determine whether there was any significant difference in mean duration between the voiceless segments of [kʰ] and [x], [k] and [x], [k] and [kʰ], as well as between the voiced segments of [g] and [ɣ], [gʱ] and [ɣ], and [gʱ] and [g].

6. DATA COLLECTION. The data for this study was taken from Hindi-Urdu language television programs from India and Pakistan, generally clips from news programs and talk shows. These were available on *YouTube* and on Indian and Pakistani news and entertainment websites. We listened to several hours of programming, scanning for Hindi-Urdu words with the target segments of /k/, /kʰ/, /x/, /g/, /gʱ/, and /ɣ/. We then saved extracts of dialog containing these segments as MP3 files, and analyzed each of them in Praat (Boersma & Weenik 2012) for duration of the target segment. For each of the six phonemes under analysis, our sample included around 25 tokens—five tokens per individual speaker—drawn from five different speakers. With 25 tokens collected for each of the six phonemes under investigation, our total sample size was 150 tokens—see data sources below the reference section for a list of audio-visual media which we used for our study.

Figures 5a through 5f give sample spectrogram and waveform data for each of the six dorsal phonemes under investigation. The spectrograms give a visual representation of particular instances of the differences in mean duration between the six phonemes. For the voiceless stops, duration was determined by measuring the period of voicelessness between preceding and following segments. For the voiced stops, duration was determined by utilizing spectral characteristics to distinguish the voiced stop from preceding and following segments.

FIGURE 5a. [k] – *kəraʧi*.

7
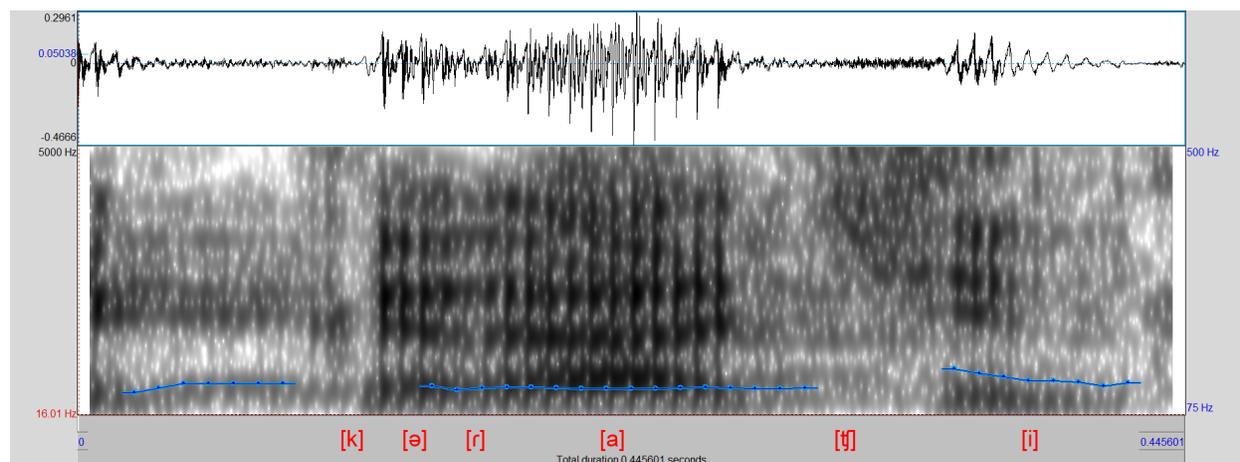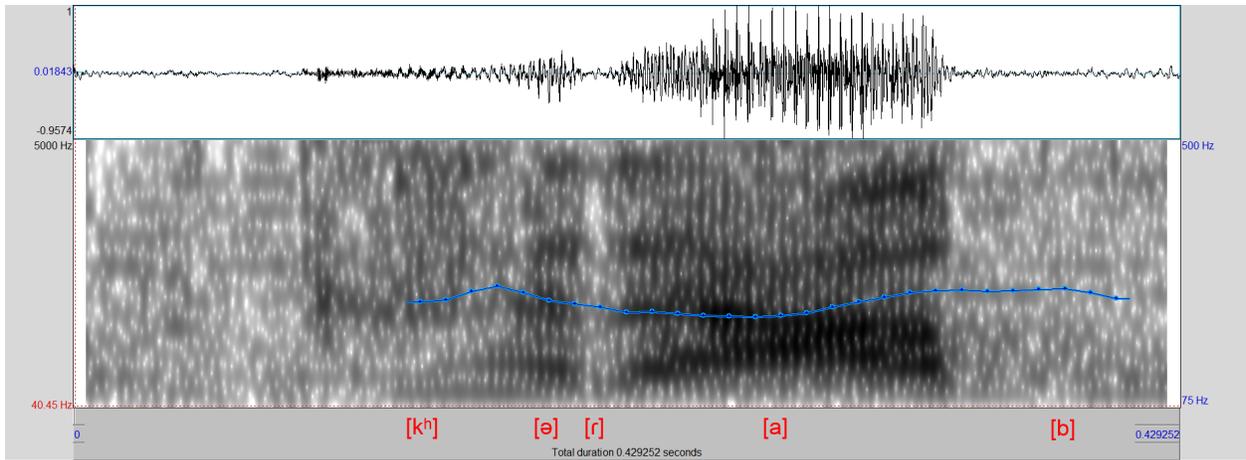
FIGURE 5b. [kʰ] – *kʰǝrab*.
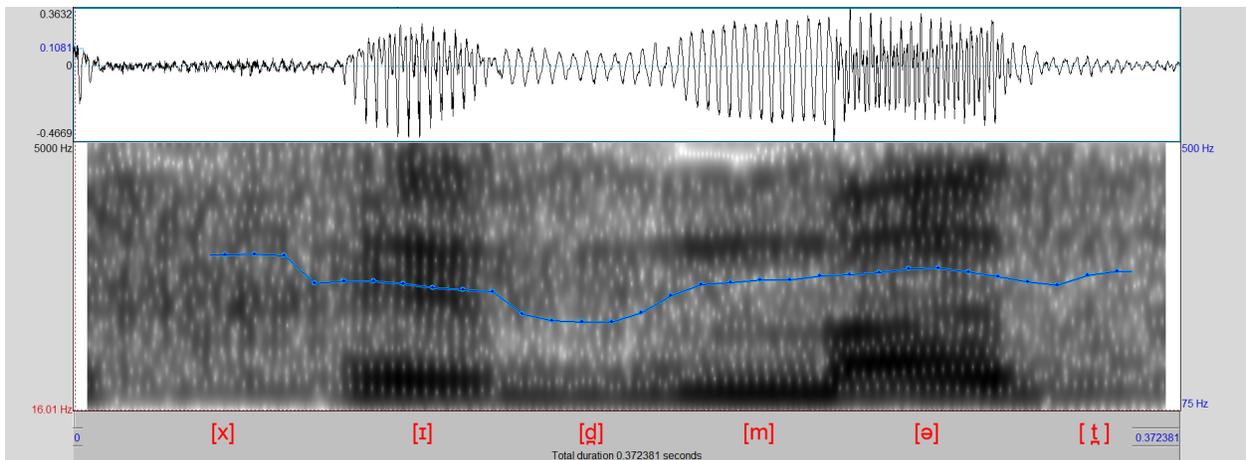
FIGURE 5c. [x] – *xid̪mat̪*.

FIGURE 5d. [g] – *gạribi*.

8

FIGURE 5e. [gʱ] – *gʱaɳʈe*.
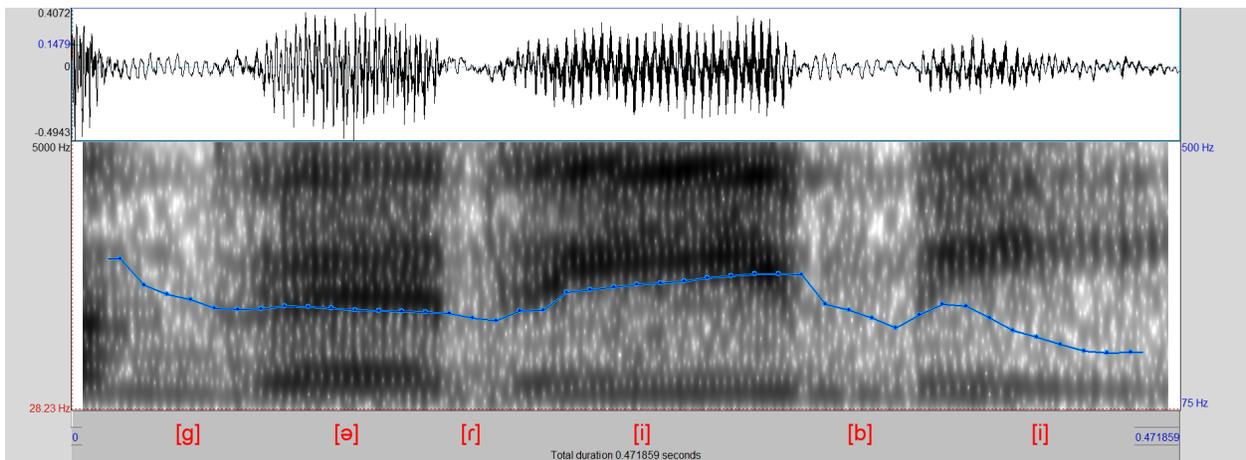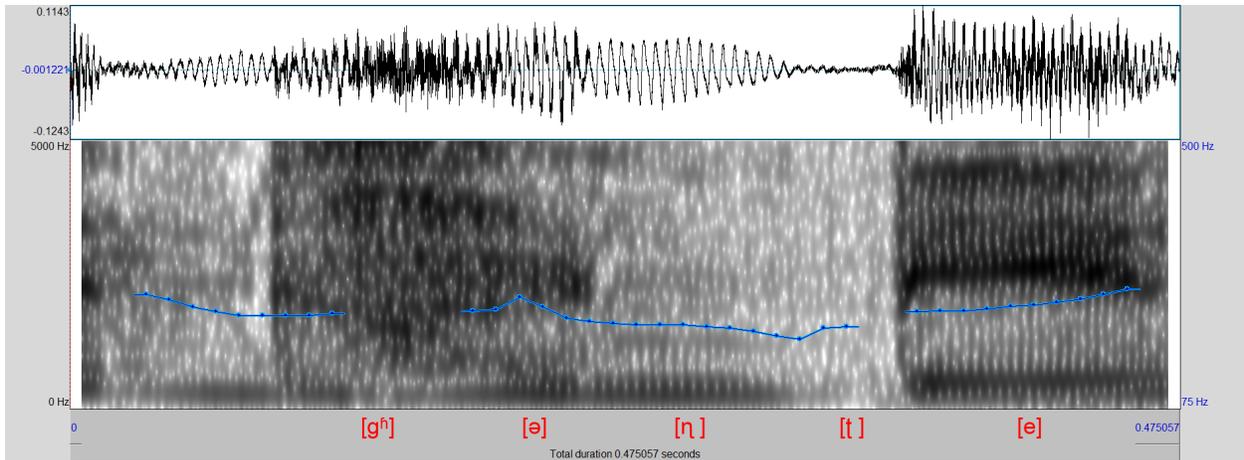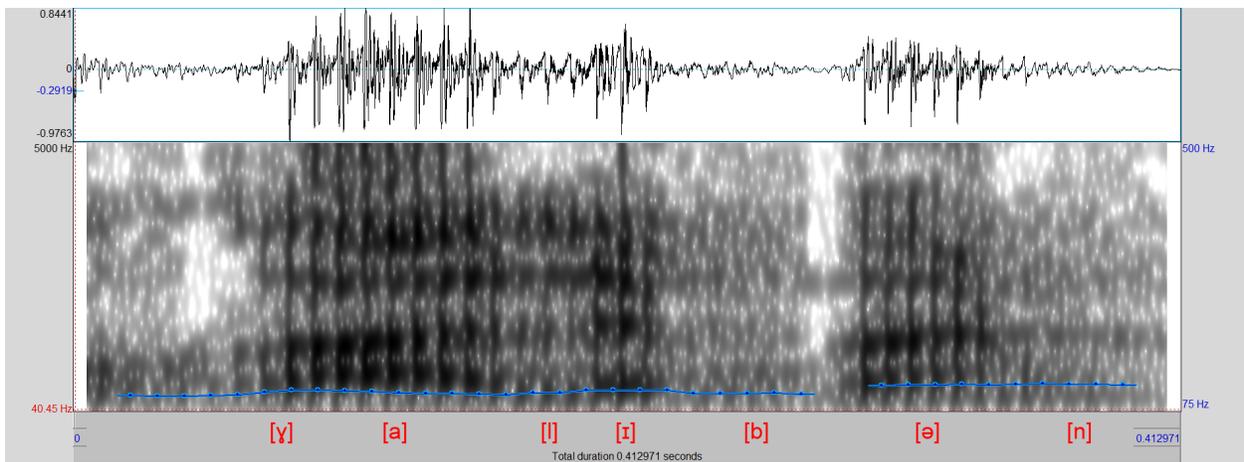
FIGURE 5f. [ɣ] – *ɣalɪbən*.

Given the preliminary scope of our project, we restricted our examination to the dorsal set /k/, /kʰ/, /x/, /g/, /gʱ/, and /ɣ/, as we have found it both practical and plausible to extrapolate our results for other places of articulation, although a more thorough investigation involving similar measurement of the other places could eventually be a useful future study.

7. RESULTS. Duration measurements for the velar series in Hindi-Urdu are given in Table 7. The data was obtained from consonant tokens taken from popular Hindi-Urdu language media, analyzed using Praat phonetic analysis software (Boersma & Weenik 2012).

	[k]	[kʰ]	[x]	[g]	[gʱ]	[ɣ]
MEAN DURATION (IN MILLISECONDS)	56	93	87	62	109	74
NUMBER OF TOKENS	33	25	26	24	25	26

TABLE 7. Mean duration of Hindi-Urdu velar consonant tokens.

9

The data in Figure 6 offers a visual representation of the relative mean durations of the six phonemes under investigation. The c-shaped bars highlight similar mean durations for the pairs [kʰ] – [x] and [g] – [ɣ].
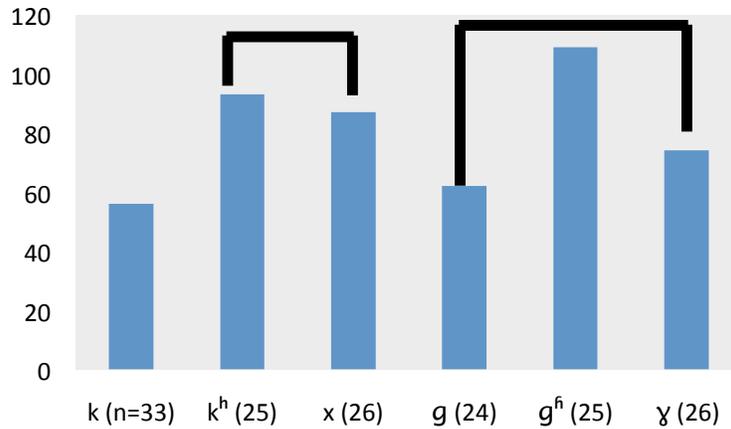
FIGURE 6. Mean duration of Hindi-Urdu velar consonant tokens (in milliseconds).

Applying a chi-squared statistical significance test, we found that there was no significant difference in mean duration between [kʰ] and [x] ($p>0.05$). In contrast, the difference in mean duration between [k] and [x], and the difference between [k] and [kʰ] were significant ($p<0.001$). There was no significant difference in mean duration between [g] and [ɣ] ($p>0.05$). In contrast, the difference in mean duration between [gʱ] and [ɣ], and the difference between [gʱ] and [g], were significant ($p<0.001$).

The data in Table 7 and Figure 6 confirm our hypothesis that the mapping of the voiceless fricatives to voiceless aspirates and the mapping of the voiced fricatives to voiced non-aspirates is due to similarities in mean duration and thus appears to be based on perception.

8. CONCLUSION. The findings of this study answer the question raised by Hock (1991:394) about the systematic nature of aspirate borrowings across languages by looking at the phonetic basis for relating fricatives to aspirate plosives. These findings, which uniquely rely on typological data supported by phonetic measurements, promise to enrich our typological understanding of the phonology of languages in contact. Most importantly, they provide a clear set of constraint distinctions to guide theories of competing motivations, such as Optimality Theory. The first way they do this is by bringing typological considerations into predicting the ranking order of constraints. The second way they do this is by strictly limiting what can be generated (in GEN) based on what a language's inventory tells us about its typology. Armed with an informed set of typological distinctions, any future work hoping to analyze subsets of this sort of data will ideally be informed by the principles outlined herein and bring us a step closer to understanding the varied yet systematic nature of languages in contact.

APPENDIX

DATA SOURCES FOR SPECTROGRAM ANALYSIS

ASAREQAYAMAT. Veena Malik angry on Mulana. *YouTube.com*. Online:
        http://www.YouTube.com/watch?v=eIO5LPjQSFw.
ASPI RE. 2010, Feb 16. Wo Kagaz Ki Kashti – Singer Samir Date – music Jagjit Singh.
        YouTube.com. Online: http://www.YouTube.com/watch?v=079ViBQ8Mro.
BBC. 2011. Pākistān mēṁ dō̄sau xufiyā qaid xānē. *YouTube.com*. Online:
        http://www.bbc.co.uk/urdu/multimedia/2011/04/110419_cage_prisoner_si.shtml.
DAWNNEWSPAKISTAN. 2010, Mar 19. News package – human smuggling. *YouTube.com*. Online:
        http://www.YouTube.com/watch?v=NIGLyD8q2nU.
DAWNNEWSPAKISTAN. 2010, Mar 25. News package – quadruplets get relief. *YouTube.com*.
        Online: http://www.YouTube.com/watch?v=jR7jGtMmVFI.
DAWNNEWSPAKISTAN. 2010, Jun 24. Lamhay – episode 2 – Part 3. *YouTube.com*. Online:
        http://www.YouTube.com/watch?v=EJYQS9-hI14.
DAWNNEWSPAKISTAN. 2010, Sep 3. News package – prices hike as Eid nears. *YouTube.com*.
        Online: http://www.YouTube.com/watch?v=QvnlFLuosK4.
DAWNNEWSPAKISTAN. 2010, Oct 14. Reporter – rich loan defaulters of a poor country – Ep61 –
        Part 4. *YouTube.com*. Online: http://www.YouTube.com/watch?v=NbOR8Ij2KKo.
DAWNNEWSPAKISTAN. 2011, Jan 21. News night with Talat- is there any relief for the masses -
        Part-3. *YouTube.com*. Online: http://www.YouTube.com/watch?v=wGaU7QjT_GI.
DAWNNEWSPAKISTAN. 2011, Feb 11. News night with Talat-Raymond. *YouTube.com*. Online:
        http://www.YouTube.com/watch?v=EN6Qxg2O7II.
DAWNNEWSPAKISTAN. 2011, Apr 17. Bolna Zaroori Hai -water crisis and domestic management
        Part 3. YouTube.com. Online: http://www.YouTube.com/watch?v=x-lSMTtfCRg.
DAWNNEWSPAKISTAN. 2011, May 3. 0305_Reporter- Pakistan caught sleeping as Osama killed –
        Ep 173-Part 3. YouTube.com. Online: http://www.YouTube.com/watch?v=sz2J2-bsgxY.
DAWNNEWSPAKISTAN. 2011, May 5. 0505_News night with Talat-Part 1. YouTube.com. Online:
        http://www.YouTube.com/watch?v=gITJG5Qfk0g.
DAWNNEWSPAKISTAN. 2011, May 5. 0505_Reporter- how did U.S evade Pakistan's air defense
        system –Ep 174- Pt1. YouTube.com. Online:
        http://www.YouTube.com/watch?v=v9tGrj1mCxQ.
DAWNNEWSPAKISTAN. 2011, May 10. 1005 Reporter Pt03. YouTube.com. Online:
        http://www.YouTube.com/watch?v=7LHSNnJhJ3s.
DAWNNEWSPAKISTAN. 2011, May 10. 1005 Reporter Pt04. YouTube.com. Online:
        http://www.YouTube.com/watch?v=kvOmWzPMTsw.
DAWNNEWSPAKISTAN. 2011, Apr 22. 2204_News night with Talat_Pt -4. YouTube.com. Online:
        http://www.YouTube.com/watch?v=OyHE1jA4QPM.
INDIA TV. 2010, May 8. India.Tv – Jk Doctor Tops. YouTube.com. Online:
        http://www.YouTube.com/watch?v=hg98MipZF0o.
INDIA TV. 2010, Jun 20. Jaya Bachchan breaks ice with Rekha. YouTube.com. Online:
        http://www.YouTube.com/watch?v=KOqCUDWwRmU.
INDIA TV. 2010, Oct 18. John Abraham in Aap Ki Adalat. YouTube.com. Online:
        http://www.YouTube.com/watch?v=a72UyMfpZQw.

INDIA TV. 2011, May 8. Farmers unrest spreads to Agra. *YouTube.com*. Online: http://www.YouTube.com/watch?v=Wacm3zYGXug.

INDIA TV. 2011, May 8. Noida toll rises to 4. *YouTube.com*. Online: http://www.YouTube.com/watch?v=HwwGGSyzNC8.

INDIA TV. 2011, May 17. India.Tv – Salman requested pretty young ladies dont call me 'BHAI'. *YouTube.com*. Online: http://www.YouTube.com/watch?v=CAD8yTAhMSI.

INDIA TV. 2011, May 17. India.Tv – Wikileaks cash. *YouTube.com*. Online: http://www.YouTube.com/watch?v=_75-eJpJ4ug.

KHOJINDIA TV. 2011, Oct 14. Khoj India – the poor & the hungry side of India. *YouTube.com*. Online: http://www.YouTube.com/watch?v=b7kqo7fsyDY.

MUSTAFA BAHTI. 2011, Jan 14. 'ajīb wa ɣarīb. *YouTube.com*. Online: http://www.YouTube.com/watch?v=aHIlFc3iHs0.

NDTV KHABAR. 2008, Nov 6. Ghar mēṁ banā hu'ā itar yā paraphyūm. *YouTube.com*. Online: http://www.YouTube.com/watch?v=JkNzjrnituI.

NDTV KHABAR. 2008, Nov 6. Ghar mēṁ banā hu'ā skin lōśan. *YouTube.com*. Online: http://www.YouTube.com/watch?v=ZSaR25n2488.

NDTV KHABAR. 2013. Ghar lauṭā ḍēṛh varṣī īśān. *khabar.ndtv.com*. Online: http://khabar.ndtv.com/LiveVideo.aspx?id=197880.

NDTV KHABAR. 2013. Lōgōṁ nē dabōcā bhāg rahē gurgē kō. *YouTube.com*. Online: http://khabar.ndtv.com/LiveVideo.aspx?id=194454.

NDTV KHABAR. 2013. Ōsāmā kē ghar kā ṭhēkēdār giraphtār. *khabar.ndtv.com*. Online: http://khabar.ndtv.com/LiveVideo.aspx?id=198576.

NDTV KHABAR. 2013. Patnī kī hatyā kā plān ḍāyarī mēṁ. *khabar.ndtv.com*. Online: http://www.ndtvkhabar.com/LiveVideo.aspx?id=165884.

NDTV KHABAR. 2013. Yē hai garībī kā paimānā.... *khabar.ndtv.com*. Online: http://khabar.ndtv.com/LiveVideo.aspx?id=199190.

RAHMANBUNAIREE. 2010, Aug 20. Indian tv debates Islam with Shahrukh Khan and Dr.Naik. *YouTube.com*. Online: http://www.YouTube.com/watch?v=Qj_KB-WpeB0.

ROBIN RAINA. 2008, Dec 23. Tv channel Voice Of India news story on Robin Raina Foundation slum homes. *YouTube.com*. Online: http://www.YouTube.com/watch?v=XteI2iYWTo4.

SHAMLAL PATEL. 2009, Mar 28. India superpower. *YouTube.com*. Online: http://www.YouTube.com/watch?v=Tmja5OwkRwc.

SHEMAROO. 2009, Mar 7. Tumhari Kasam – Hum Dono Milke Kaagaj Ke Dil Ke – Mukesh – Asha. *YouTube.com*. Online: https://www.youtube.com/watch?v=IGQY-SAl0-4.

ZOOMDEKHO. 2008, May 19. When Shahrukh Khan got angry. *YouTube.com*. Online: http://www.YouTube.com/watch?v=3586n0CD9PU.

ZOOMDEKHO. 2009, Mar 31. Salman Khan Katrina Kaif break up. *YouTube.com*. Online: http://www.YouTube.com/watch?v=zaH9_CPRqMs.

ZOOMDEKHO. 2011, Feb 2. Planet Bollywood – Salman's new 'partner', Kapoors sort the confusion and more news from Bollywood.... *YouTube.com*. Online: http://www.YouTube.com/watch?v=0dzeoOM9ykY.

REFERENCES

ADLER, ALLISON N. 2006. Faithfulness and perception in loanword adaptation: A case study from Hawaiian. *Lingua* 116.1024–1045.

AHN, SANG CHEOL. 2003. English interdental substitution. *University of Helsinki*. Online: http://www.ling.helsinki.fi/kielitiede/20scl/abstracts.shtml#Ahn.

ASAHI SHIMBUN COMPANY. 2005. Tsukame bokuranonatsu 77-kō no taisen kimaru. *Asahi.com*, 03 July 2005. Online: http://www.asahi.com/koshien/localnews/TKY200507030090.html.

BBC. 2002. 'Guru' kō bhārī saphalatā. *BBC Hindi*, 29 August 2002. Online: http://www.bbc.co.uk/hindi/news/020828_guru_sz.shtml.

BOERSMA, PAUL, and DAVID WEENIK. 2012. Praat: Doing phonetics by computer [computer program]. Online: http://www.praat.org/.

COWELL, MARK W. 2005. *A reference grammar of Syrian Arabic (based on the dialect of Damascus)*. Washington, DC: Georgetown University Press.

DINNSEN, DANIEL A., and JUDITH A. GIERUT. 2009. Optimality theory: A clinical perspective. *The handbook of clinical linguistics*, ed. by Martin J. Ball, Michael R. Perkins, Nicole Müller, and Sara Howard, 439–451. Malden, MA: Blackwell Publishing.

DINNSEN, DANIEL A., and LAURA W. MCGARRITY. 1999. Variation and emerging faithfulness in phonological acquisition. *Twenty-third annual Boston University conference on language development*, ed. by Annabel Greenhill, Heather Littlefield, and Cheryl Tano, 172–183. Somerville, MA: Cascadilla Press.

EISHUPPAN-SHA. 2010. *3-Shūkan de hara o hekoma seru 7-tsu no tekunikku*. Eishuppan-sha.

ELBERT, SAMUEL H. 1964. Hawaiian reflexes of Proto-Malayo-Polynesian and Proto-Polynesian reconstructed forms. *The journal of the Polynesian society* 73.399–410.

GENETTI, CAROL. 2007. *A Grammar of Dolakha Newar*. Berlin: Walter de Gruyter.

GONET, WIKTOR, and GRZEGORZ PIETROŃ. 2006. English interdental fricatives in the speech of Polish learners of English. *Neofilologia* 8.73–93.

HOCK, HANS HENRICH. 1991. *Principles of historical linguistics*. Berlin, Germany: Walter de Gruyter.

KIM, HYUNSOON. 2009. Korean adaptation of English affricates and fricatives in a feature-driven model of loanword adaptation. *Loan phonology*, ed. by Andrea Calabrese and Leo Wetzels, 155–180. Philadelphia, PA: John Benjamins Publishing Company.

KŌREI-SHA JŪKANKYŌ KENKYŪJO. 2009. *Bariafurī jūtaku dokuhon*. Sanwa shoseki.

KUMAR, R. 2011. Ghātak mithēn kē duṣprarōkanē kī anūṭhī pahal. *State Hindi news*, 13 December 2011. Online: http://statehindinews.blogspot.com/2011/09/blog-post_7153.html.

LABRUNE, LAURENCE. 2012. *The phonology of Japanese*. New York: Oxford University Press.

MADDIESON, IAN. 1984. *Patterns of sounds*. New York, NY: Cambridge University Press.

MAEDA KYOKO OFFICE. 2007. *Ki ni naru! Ren'ai shinri tesuto*. Saitō sha.

NAV BHARAT TIMES. 2008. Hēdar kī harakat nē hairat mēṁ ḍālā. *Navbhārat ṭā'im*. Online: http://navbharattimes.indiatimes.com/articleshow/2879193.cms.

PARADIS, CAROLE, and DARLENE LACHARITÉ. 2012. The influence of attitude on the treatment of interdentals in loanwords: Ill-performed importations. *Catalan journal of linguistics* 11.97–126.

PARK, HANYONG. 2007. Varied adaptation patterns of English stops and fricatives in Korean loanwords: The influence of the P-map. *Indiana University Linguistics Club Working Papers* 7. Online: https://www.indiana.edu/~iulcwp/contents.cgi?which=7.

PAYNE, JOHN R. 2009. Persian. *The World's Major Languages* (2nd ed.), ed. by Bernard Comrie, 514–522. New York: Routledge.

RĪPIN, RIZĀZU. 2000. *Patān katsuyō jiko shōkai no eigo handobukku*. Sōgen-sha.

SAKODA, KENT, and JEFF SIEGEL. 2004. Hawai'i Creole: Phonology. *A handbook of varieties of English, volume 1: Phonology*, ed. by Edgar W. Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie, and Clive Upton, 729–749. Berlin: Mouton de Gruyter.

TAMAOYŌICHI. 2009. *Adobe Illustrator CS 4*. Shūwa shisutemu.

WEINBERGER, STEVEN H. 1997. Minimal segments in second language phonology. *Second-language speech: Structure and process*, ed. by Allan James and Jonathan Leather, 263–311. Berlin: Walter de Gruyter.

NOTES

[1] Common manner-faithful borrowings of English words into Japanese include instances of [θ] to [s], such as *heru*s*um*ē*t*ā < *heal*th*meter* (Eishuppan-sha 2010:180), *yunittoba*s*u* < *unit ba*th (Kō rei-sha 2009:120), s*ur*ī*saizu* < th*ree size* (i.e. of standard body measurements) (Rīpin 2000:51), and instances of [ð] to [z] such as *sum*ū*z*u < *smoo*th (Tamaoyōichi 2009:98), *ma*z*akon* < *mo*th*er com[plex]* (Maeda 2007:96), *iy*ā *obu* z*a k*ō*chi* < *year of* th*e coach* (Asahi 2005:par. 2). Well-known L2 descriptive works also agree that Japanese interdental fricatives [θ] and [ð] are borrowed manner-faithfully as [s] and [z] (Labrune 2012:98–99) and (Weinberger 1997:275–276).

[2] Typically, the Persian voiceless stops are phonetically aspirated, but there is no phonemic aspiration contrast.

[3] From Kumar (2011).

[4] From BBC (2002) and Nav Bhārat (2008).

Easier Said Than Done:
Metaphor Interpretation between English and American Sign Language

Benjamin Anible
University of New Mexico

ABSTRACT. A pilot study performed on interpreted discourse of the 2008 presidential primaries found that when English-to-ASL interpreters encounter idioms on different ends of the continuums of CONVENTIONALITY and COMPOSITIONALITY, the type of interpretation they give is predictable. Shared metaphors received a higher proportion of transcoding interpretations than metaphors that were not shared. Additionally, opaque and fixed metaphors received more conceptual translations then other type of idioms. These results paint a picture of the unique challenge that encountering metaphors can pose while engaged in simultaneous interpretation and shed light on the nature of ASL metaphors.

1. INTRODUCTION. Expressing the intricacies of human thought is beyond the capabilities of literal language alone. Abstract concepts, scientific theories, the vagaries of emotion, and even many mundane daily activities are best conveyed through the use of metaphor. Steven Colbert once jokingly asked Elizabeth Alexander, the author of the poem 'Praise Song for the Day' (read at the inauguration of President Barack Obama), 'what's the difference between a metaphor and a lie?' to which she replied 'the two are not necessarily exclusive, metaphor ... is how you use the language to *increase* meaning' (Colbert 2009). Metaphor is often (and occasionally humorously) misconstrued as a special exception to normal language use.

There are many motivations for the use of metaphor in discourse. This paper discusses the psychological structures that make metaphor possible, reviews how metaphor has traditionally been viewed, explores specific metaphorical mechanisms in spoken and signed discourse, and considers how metaphor can affect bimodal interpretation (interpretation between a signed and spoken language).

2. CONCEPTUAL METAPHOR. The example from the Colbert interview above shows that metaphor might be considered a special case of language use. However, some linguists (e.g. Lakoff & Johnson 1980, Fauconnier & Turner 2002) have come to believe its influence is deeply rooted in our physical experience. This idea is commonly referred to as CONCEPTUAL METAPHOR THEORY. At the core of this theory is the idea that the way we think is inherently metaphorical, because we generalize from our interaction with the world to a higher order of abstraction. A conceptual metaphor is a quick and dirty way we interact with things in the world that are beyond concrete experiential perception. A good example of this is *anger*. Anger is not something we can see or taste or hear or touch. When we talk about anger, we use a metaphor; a relationship in our mind between a concrete thing and an abstract thing. See 1 below.

(1) Bill really blew his stack today in class.

The phase 'blew his stack' is an expression of a conceptual metaphor along the lines of Anger Is A Heated Fluid In A Container. Here, a 'stack' is conceptualized as something that can

undergo a violent change in state ('blowing') due to the heating of a fluid. The expression likely comes from the flap on steamboat smoke stacks.

Lakoff and Johnson's (1980) theory of conceptual metaphor convinced many linguists about the primacy of metaphor in thought and language. In the 2003 afterword to *Metaphors We Live By*, four fallacies that have proved to be stumbling blocks for the acceptance of conceptual metaphor theory are outlined: metaphor is merely a matter of words, metaphor is based on similarity, concepts are literal and cannot be metaphorical, and rational thought is not shaped by the nature of our physical bodies. Lakoff and Johnson argue that, in fact, metaphor is concept-based, not word-based. Metaphor is not based on similarity, but on cross-domain correlation. Even basic concepts are reasoned about via multiple metaphors; the concept of time is often mapped on to our understanding of spatial relations. Metaphor is inherent in the way our bodies interact with the world daily.

Lakoff and Johnson also explain that there is a neural basis for metaphorical thought. Primary metaphors (those that more complex metaphors build on) arise via CONFLATION Our earliest experiences of affection, or being held, create warmth, which gives the metaphor Warmth Is Affection. Neural activation occurs in the regions associated with temperature and those associated with emotion. As 'neurons that fire together wire together', this physical construction of two sets of neurons *is* the structure of the metaphor itself (Lakoff & Johnson 1980).

Traditionally, a metaphor is defined as the 'transfer of relations from the semantic field of the vehicle to the conceptual domain of the topic' (Wilcox 2000:8), where a vehicle is the focal point in a metaphor (also indicating the aspect of 'transport') and the topic is the subject of the transfer. In constructing metaphorical meaning, the vehicle has a privileged position. The comparison between the two domains is asymmetrical. Anecdotally people always talk about how all metaphors break down at some point, which appears to take effect when a bidirectional comparison is made.

Metaphor has also traditionally been considered one type of TROPE, or figure of speech. Important tropes for understanding metaphor in signed languages are simile, metonymy, synecdoche, and experiential metaphor. Synecdoche has been called a master trope in which aspects of both metaphor and metonymy take effect. Metonymy is a where a part of a domain stands for the whole. Since metaphor is when one domain stands for another, a synecdoche is when part of one domain stands for part of another. The difference between simile and metaphor is the subject of extensive debate but can be schematically thought of as a qualitative rather than a quantitative distinction (Wilcox 2000). Finally experiential metaphors are those that are deeply ingrained into our everyday language use and include ORIENTATIONAL, ONTOLOGICAL and STRUCTURAL distinctions. Wilcox also notes that metaphorical extensions are a critical component of language change and that it is a device for adapting language to culturally specific communicative needs.

In spoken discourse, it has been recognized that metaphor can be used for a broad range of communicative intentions. Cacciari (1998) proposes five main reasons speakers elect to employ metaphorical expression. The first is to bridge the gap between abstract mental conceptualizations and perceptual experiences. A good example is the use of color terms. Using the word *white* to mean honest or clean rather than the word *purple* is not just an artifact of the language. Rather, the properties of white things in our perceptual experience (such as a field of snow) carry a quality of the mental realities we use *white* to explain.

The second reason for using metaphor is to express emotional experience. Mac Cormac

(1985) elaborates on this idea, suggesting there is an emotive imperative in metaphor that creates a bond between interlocutors: 'The illocutionary force of sharing may constrain the variety of perlocutions that occur' (161). What he means is that the performative act of uttering a metaphor affects a particular reaction in a listener that a purely literal utterance of an equivalent idea would not necessitate.

Another reason Cacciari outlines, is to set or change the conceptual perspective of the speaker and their audience. Metaphor evokes cultural models. Metaphors can also force us to see a situation or idea from a different perspective than we originally had been viewing it in. Also, metaphors are used for politeness and saving face. Figurative language can be ignored or misunderstood—segmenting an audience into those who get it and those who don't. Finally, metaphor is used to summarize bundles of proprieties. Metaphors' interpretations depend on the fact that they provide a compact form of expression for complex ideas. 'A literal paraphrase of a metaphor inevitably says too much—and with the wrong emphasis' (143). Cacciari finally ruminates that metaphors are epistemologically and communicatively necessary. Metaphor is an iconic tool of language just as the hammer is the iconic tool of the arm and the telescope the iconic tool of the eye.

The discourse use of metaphor in spoken language proposes few constraints (though clearly there are *patterns*) on the specific phonological and lexical forms that metaphor can take for particular communicative intent. Turning now to the way that signers employ metaphor, we will find that there are unique restrictions that spring from the visual modality, namely ICONICITY.

3. ICONICITY. American Sign Language (ASL) linguistics has had a rocky relationship with the topic of iconicity. Because iconicity and metaphor are so intertwined with each other, an understanding of the evolving perspective on iconicity is important to explore. Briefly, arbitrariness was famously proposed to be a prerequisite to recursivity (Saussure 1916), and by extension to be a prerequisite to human language (Hockett 1960). Early sign linguists needed to obscure the role of non-arbitrary signals to cement the legitimacy of signed languages. Interestingly, iconicity in the aural modality has also been recognized. At the discourse level, events in a narrative are generally recited in chronological order (Haiman 1980). At the sentence level, constituents that are more relevant to a given verb are syntactically close to it (Bybee 1985). At the morpho-phonemic level, phonesethemes like *glisten, gleam*, *glitter* have restricted but regular form-meaning (gl=bright) correlations. At the phonemic level some sound patterns in English are statistically associated with concreteness and imageability (Reilly & Kean 2007).

A number of sign language researchers, however, have recognized the centrality of iconicity to signed languages. Many, unfortunately, mistook iconic relationships as metaphoric. All the errors stem from recognition of a mapping, but failing to disambiguate form-meaning connections from form-meaning mappings with an additional meaning-meaning facet. Mandel (1977) recognized metonymic mappings like the profiling of 'beard' in OLD, but misidentified the purely iconic mappings of the 'branches' in TREE as metaphoric. Boyes-Braem (1981) showed ASL is iconic while still allowing recursivity by virtue of its regularity. Gee and Kegel (1982) discuss how the concrete experience of locative manipulation is iconically productive in ASL in such a way as to allow the infinitely recursive expansion of location components to novel lexical items. Wilbur (1987) is a notable, but cursory, exception in that she recognized the *double* mapping for a sign like BRILLIANT where SEEING IS UNDERSTANDING is interleaved with the concept of shining and in turn, iconically mapped to the 8 handshape and a sparkle-like

motion. Brennan (1990), analyzing BSL, makes the same error as Boyes-Braem in calling iconic signs metaphorical.

Taub (2001) proposes a model for the creation of iconic forms in both signed and spoken language which she calls the ANALOGUE BUILDING model. The model consists of three stages: image selection, schematization and encoding. In the first stage, an image is selected from the store of concepts for a particular individual. This is an inherently metonymic process since the total meaning of all attributes for a particular concept is always more than can be represented by a particular symbol and some parts must be focused. In the second stage, the selected properties are abstracted and generalized such that only the properties that are important for a specific language are retained. In the final stage of the process the selected schema is encoded into actual linguistic form. Encoding is necessarily arbitrary, because any given language has a set of tools (for signed languages a set of morphophonemic parameters) that are conventionalized in their use.

Metaphors are incorporated into the model by target-source correspondences made before image selection. Additionally, metaphorical signs are pre-schematized, meaning the relevant concepts have already been selected by the metaphor mapping process. For the encoding stage, there are no differences between iconic and metaphorical-iconic signs.

Meir (2010) continues the examination of the interaction between iconicity and metaphor in sign languages looking at Israeli Sign Language (ISL). Following Taub, she defines iconicity as a set of selected correspondences between the form and meaning of a sign. This allows for some signs to be fully iconic or partially iconic, given the number of correspondences between all the possible parameters of form and meaning that exist for the sign in question. Meir notes, however, that there is a constraint on the possible metaphorical mappings of iconic forms. She calls this the 'Double-Mapping Constraint (DMC): A metaphorical mapping of an iconic form should preserve the structural correspondences of the iconic mapping. Double mapping should be structure-preserving' (879).

An example of this is the ISL sign for FLY, which iconically maps arm flapping to wing flapping. This form resists metaphoric collocations like 'time flies' that exist in many spoken languages because the source 'wing flapping' does not meaningfully map to the target 'time'.

While ECHOICS in the spoken modality do not seem to be constricted by the DMC, idiomatic expressions do share similarities to the observable visual modality limitations. The contexts in which an idiom may be used have to be compatible with the mental image invoked by the literal meaning. For example, the idiom 'to be caught between a rock and a hard place', which roughly means 'to be in a very difficult position; to face a hard decision', can be used only in contexts in which a protagonist is being trapped between two very difficult obstacles, and not in any situation where the protagonist faces a hard decision (ibid).

The iconic blocking of free mapping in signed language metaphor might also account for the interesting fact that one of the more common ways spoken languages express a conceptual metaphor—the idiom—is very infrequently observed. Before unpacking this idea, first consider the role of idiomatic expressions for conveying conceptual metaphor in spoken language.

Gibbs et al. (1997) set out to answer the question: 'Do people quickly access conceptual metaphors each time an idiom is encountered in discourse?' The presupposition behind this question is that metaphor is more than a linguistic artifact, but rather a significant aspect of cognition. Four different hypotheses are offered as possible answers to the question.

H1: Metaphoric impact can be traced diachronically in expressions.
H2: Metaphor as understood in a community motivates interpretation of expressions.
H3: Metaphor as understood by an individual motivates interpretation of expressions.
H4: Metaphor is an online processing phenomenon.

To test these hypotheses, the author ran two experiments. The first was aimed to test whether metaphorical knowledge is accessed in online processing. This was done in a lexical decision–priming paradigm where the comprehension time of a conceptually related word to a previously displayed metaphor phrase was measured. Expression 2a was presented, after which subjects reported whether either a conceptually related word, 2b, or conceptually not related word, 2c, was a lexical item in English.

(2) a. cross that bridge later
     b. water
     c. waste
     d. fix that problem later

Results indicated that lexical decision times were in fact lower in the test scenario 2b, indicating positive evidence for hypothesis H4. The average RT for a related word was 899 ms (s.d. .05), but 914 ms (s.d. .03) for an unrelated one. An alternative interpretation of these results is that the core concept in the metaphor and the literal interpretation of the idiom were in alignment somehow. A control test used literal paraphrases, 2d, against idioms and still found response times in a conceptually related lexical decision task faster after reading the idiom than the literal paraphrase. The faster RT for related items suggests activation of specific conceptual knowledge.

Gibbs et al.'s experiment presents convincing evidence that idioms like those in 1 and 2a activate underlying conceptual metaphors like Anger Is A Heated Fluid In A Container or Problems Are Bodies Of Water. If a process like this exists in spoken languages like English is there an equivalent structure in a signed language like ASL? There are some examples of collocations that express a conceptual metaphor in ASL such as 3 and 4, but they are much less common than they are in English. In 3, the meaning 'you're too late' is an expression of a metaphor along the lines of Conversations Are Entities. In 4, the meaning 'I've already been there' is an expression of Physical Contact Is Being In A Location, where the touching of the hand is the physical contact. It may be that even for these examples, the only cross-domain mapping occurs with the signs TRAIN and TOUCH respectively and the other materials in the collocations do not contribute to the meaning of the core lexical items of the metaphor itself.

(3) TRAIN GO SORRY
(4) TOUCH FINISH

Why then, are there so few examples of ASL metaphorical collocations (idioms)? One possible explanation, as mentioned above, is that restriction of the double mapping constraint doesn't allow for additional signs to contribute to the meaning of the metaphor as a whole. Many signs bring their own iconic construal (following Taub's model) of conceptual content that interferes with the very specific construal the conceptual mapping selects. In this way there simply is no need for a collocation. Single metaphorical signs are able to perform the same

functions, using the same cognitive processes that idiomatic expressions do in English. Another possible explanation is that ASL and other signed languages around the world are not old enough to have the sort of idiomatic artifacts that are so readily apparent in English. Indeed, linguists have recognized that there is a tendency in ASL for iconic forms to become less iconic (Frishberg, 1975). It is conceivable that as forms become less and less iconic, more metaphorical collocations may arise.

4. PILOT STUDY. Having recognized that there are two unique ways the forms of English and ASL metaphorical expressions are realized (single vs. multi-word instantiations), it is potentially enlightening to consider how interpreters working between these two languages navigate the difference. Do restrictions on form-meaning mappings in English affect how idioms are interpreted into ASL, where such mappings interact with purely metaphorical ones?

First, it is useful to categorize English idioms as a function of diachronic semantic shift that varies on two continuums as first elaborated by Fillmore, Kay, and O'Connor (1988). These continuums are perhaps most easily thought of as CONVENTIONALITY and COMPOSITIONALITY. On one end of the conventionality scale there are extremely productive forms like the form 'the X-er the Y-er', where any number of lexical items can be substituted for X and Y, as in 5a and 5b. On the other end, there are extremely fixed forms that cannot have any of their constituents substituted without losing their meaning like 'give it up'— meaning applause. The other continuum of compositionality has very transparent idioms on one end like 'wide awake', where the meaning of each word contributes nearly fully to the meaning of the phrase as a whole. Opaque idioms, by comparison, are not immediately understandable from the literal meaning of the words that substantiate them (consider 'a snake in the grass'). See Figure 1 for an illustration of how example metaphors fall out along the two continuums.

(5) a. the bigger the better
b. the smaller the brain the smarter you are

A pilot study performed on interpreted discourse of the 2008 presidential primaries found that when English-to-ASL interpreters encounter idioms on different ends of these continuums, the type of interpretation they give is predictable. Triandis (1976), following in the footsteps of a long tradition of literary translation, identifies that interpretations themselves vary on a formal/dynamic continuum of equivalence. Formal equivalence focuses attention on the message whereas dynamic equivalence focuses attention on the intended audience.

Using this distinction as a starting point, three types of interpretation were selected that incorporate the stages of metaphor construction in sign language while simultaneously incorporating the distinction between formal and dynamic equivalence. Twenty-five tokens were obtained from YouTube over a series of four video-clips by random selection. The selection of tokens from a specific register and genre was carried out in an attempt to factor out as much variation as possible given the small sample size. All coding of English idiom type and translation type were performed by the author.

Compositionality
Transparent

productive transparent "...the still waters of peace..."	**fixed transparent** "...lower our sights..."

Productive ——————————————————— *Fixed*

productive opaque	**fixed opaque** "...give it up..."

Conventionality (vertical axis label)

Opaque
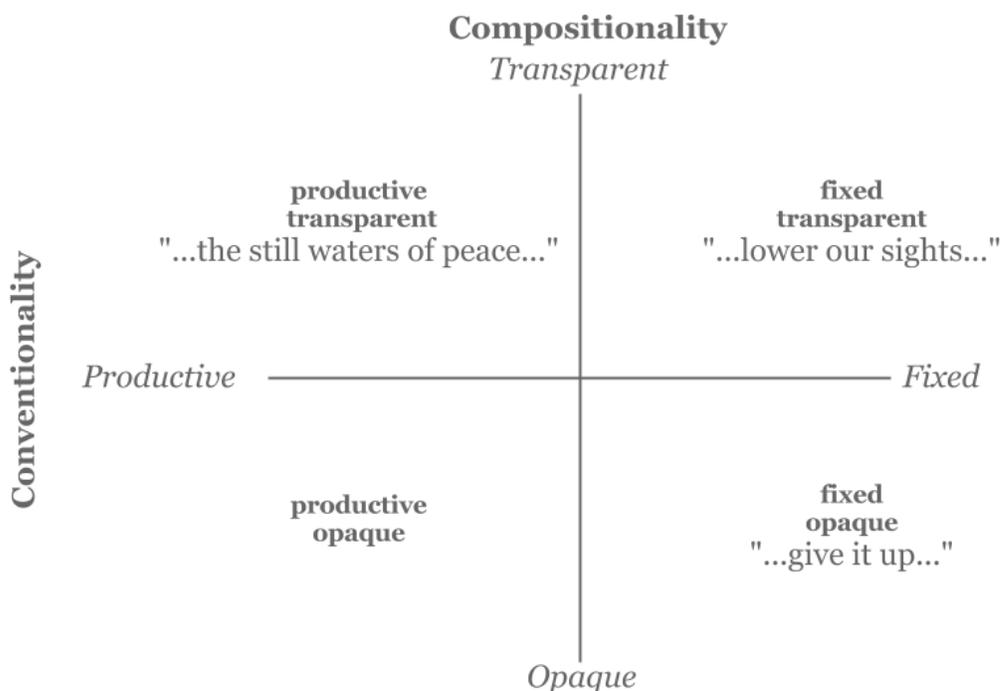
FIGURE 1. Types of metaphors on the scales of compositionality and conventionality.

The first type of interpretation—TRANSCODING, as in 6, neither renders equivalence of the metaphor's image selection, nor does it give equivalence in the encoding. In other words, while in English *met* in this context conveys the concept of confrontation, in ASL the sign MEET does not carry this connotation and the iconic mapping of the sign does not include this connotation either. It is attempting to convey formal equivalence but fails. The second type of interpretation—CONCEPTUAL, as in 7, conveys the dynamic equivalence in the image selection but fails to give formal equivalence in encoding. The third type—DEPICTIVE, as in 8, conveys both formal and dynamic equivalence at all stages of the metaphor creation process of the analogical building model.

(6) These challenges will be **met**. → CHALLENGE **MEET** WILL.
(7) These are the **voices** of Americans. → PRO.3 **VOICE** AMERICA
(8) I am not running for president to put a **band-aid on our problems**. →
    PRO.1 NOT RUN FOR PRESIDENT **CL:'put on band-aid'** PRO.2-poss PROBLEM

A possible predicting factor in the study was whether or not both English and ASL have metaphors that select the same image from the set of possible concepts that can be used to describe a given event. 9a is an example of a metaphor that is shared between ASL and English and 9b is an example of one that is not. 9b must be interpreted literally.

(9) a. he blew his stack → BOIL-WITH-ANGER
    b. cross that bridge when we come to it → THAT PROBLEM RESOLVE LATER

A Fisher's Exact test showed no significant effect of sharedness on the type of interpretation ($p = .31$), though there was a trend for more depictive interpretations for the non-shared category than other categories. Conceptual interpretations were the least likely if the metaphor was not shared. No significant effect of conventionality on the type of interpretation ($p = 0.11$) was found. Fixed conventionality resulted in a high chance of conceptual interpretation. There was a significant effect of compositionality on the type of interpretation ($p < .05$). Opaque compositionality resulted in a significantly higher chance of a conceptual interpretation. See Figure 2 for results.

Compositionality
Transparent

$n = 5$	Trans	Con	Dep
shared	1	0	0
not	2	0	2

$n = 3$	Trans	Con	Dep
shared	0	0	1
not	1	0	1

Productive —————————————————— *Fixed*

Conventionality

$n = 17$	Trans	Con	Dep
shared	2	2	0
not	4	6	3

Opaque
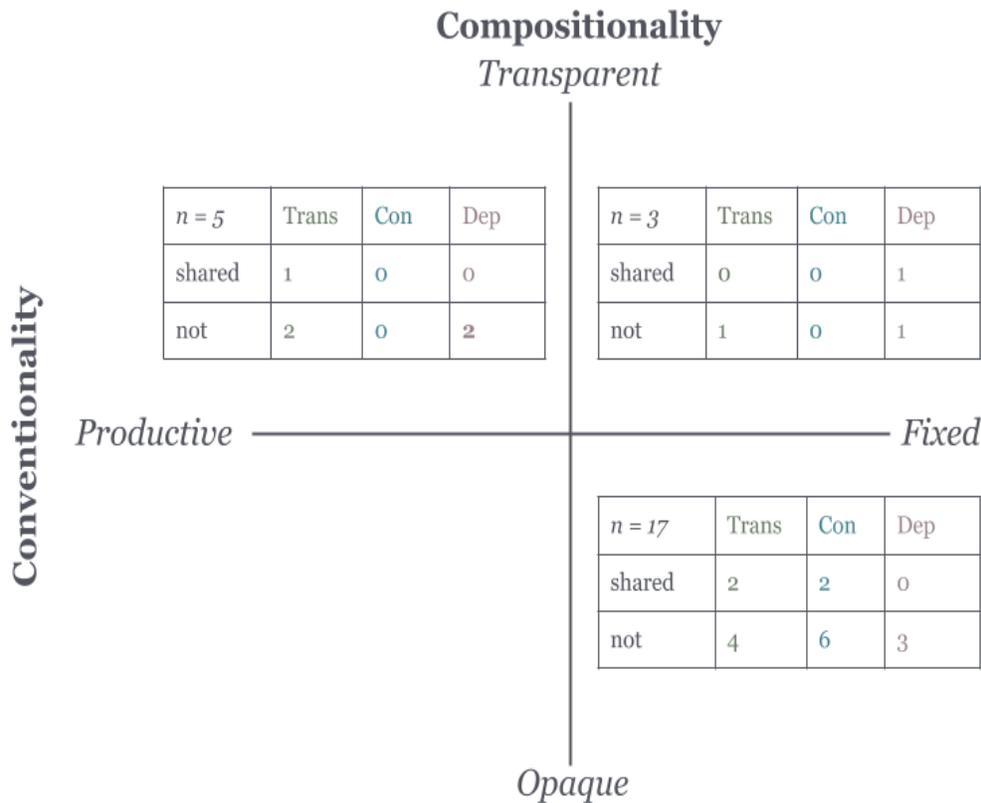
FIGURE 2. Types of interpretation categorized by type of metaphor.

While not significant in this sample, shared metaphors received a higher proportion of transcoding interpretations than metaphors that were not shared. Since depicting and conceptual interpretations more directly necessitate activation of semantic information, these types were used when metaphors were not shared. Additionally, opaque and fixed metaphors received more conceptual translations than other types of idioms. Together these results paint a picture of the unique challenge that encountering metaphors can pose while engaged in simultaneous interpretation between signed and spoken languages. The role the iconic component of signed language metaphor creation adds should not be discounted when considering this challenge.

5. CONCLUSION. This paper has explored how a conceptually motivated understanding of metaphor allows for the widespread and daily use of metaphor in both thought and language. It

should not be surprising that signed languages also have the potential for rich and pervasive metaphor use. Linguists now recognize that a theory of signed language metaphor needs to account for iconic mappings.

The results of the pilot study described in this paper provide quantitative data to support this understanding of metaphor in ASL. There is no reason for depictive translations, which are most similar to well-formed visual modality metaphors in the theory proposed, to be preferred when the English text is using the least transparently metaphorical metaphors (fixed and opaque) unless source domains are conflicting with iconic motivations. In this case there is a smaller risk of the kind of iconic blocking proposed by Meir (2010). Put another way, when an interpreter is aware of the source domain of an English metaphor they are constrained in the number of form–source mappings they can employ in their ASL translation since some or all of those mappings would result in a nonsensical metaphorical extension. Comparatively, when an interpreter is unaware of an English metaphor's source domain motivation, they have more freedom to render an ASL interpretation with double mappings. Such interpretations are certainly preferable as they allow for functional equivalence in addition to conceptual equivalence.

It is reasonable to assume that the interpreters in this study were, indeed, aiming to achieve functional equivalence when possible. This is because when metaphors were shared between English and ASL transcoding interpretations we preferred. These transfer the English source–target domain mappings that are so important directly into ASL. While consumers of this kind of interpretation need to access their own knowledge of English metaphors to achieve comprehension, the import of the source text selection of a metaphor is still conveyed.

Future studies that can accommodate more rigorous controls on interpreter proficiency and metaphor selection are anticipated to more clearly reveal the trends observed here.

REFERENCES

BOYES-BRAEM, PENNY. 1981. *Features of the handshape in American Sign Language*. Berkeley: University of California dissertation.

BRENNAN, MARY. 1990. *Word formation in British Sign Language*. Stockholm: University of Stockholm dissertation.

BYBEE, JOAN L. 1985. Diagrammatic iconicity in stem inflection relations. *Iconicity in syntax,* ed. by John Haiman, 11–47. Amsterdam: John Benjamins.

CACCIARI, CHRISTINE. 1998. Why do we speak metaphorically: Reflections on the functions of metaphor in discourse and reasoning. *Figurative Language and Thought*, ed. by Albert N., Christine Caccari, Raymond W. Gibbs and Mark Turner, 119–57. New York: Oxford University Press.

COLBERT, STEVEN. 2009, Jan. 21. *Elizabeth Alexander explains to Stephen the difference between a metaphor and a lie* [Television broadcast]. New York: Comedy Central

FAUCONNIER, GILLES, and MARK TURNER. 2002. *The way we think: Conceptual blending and the mind's hidden complexities.* New York: Basic Books.

FILLMORE, CHARLES J.; PAUL KAY; and MARY CATHERINE O'CONNOR. 1988. Regularity and idiomacity in grammatical constructions: The case of 'let alone'. *Language* 64.501–38.

FRISHBERG, NANCY. 1975. Arbitrariness and iconicity: Historical change in American Sign Language. *Language* 51.696–719.

GIBBS, RAYMOND W. JR.; JOSEPHINE BOGDONOVICH; JEFFREY R. SYKES; and DALE J. BARR. 1997. Metaphor in idiom comprehension. *Journal of Memory and Language* 37.141–54.

GEE, JAMES P., and JUDY A. KEGL. 1982. Semantic perspicuity and the locative hypothesis: Implication for acquisition. *Journal of Education* 3.185–209.

HAIMAN, JOHN. 1980. The Iconicity of grammar: Isomorphism and motivation. *Language* 56. 515–40.

HOCKETT, CHARLES F. 1960. The origin of speech. *Scientific American* 203.89–97.

LAKOFF, GEORGE and MARK JOHNSON. 1980. *Metaphors we live by*. Chicago, Ill.: University of Chicago Press.

MAC CORMAC, EARL R. 1985. *A cognitive theory of metaphor.* Cambridge, Mass.: MIT Press.

MANDEL, MARK. 1977. Iconic devices in American Sign Language. *On the other hand: New perspectives on American Sign Language*, ed. by Lynn A. Friedman, 57–107. New York: Academy Press.

MEIR, IRIT. 2010. Iconicity and metaphor: Constraints on metaphorical extension of iconic forms. *Language* 86.865–96.

REILLY, JAMIE and JACOB KEAN. 2007. Formal distinctiveness of high and low-imageability nouns: Analyses and theoretical implications. *Cognitive Science* 31.157–68.

SAUSSURE, FERDINAND DE. (1916). *Cours de linguistique generate [Course in general linguistics]*. Paris: Payot.

TAUB, SARAH. 2001. *Language from the body*. New York: Cambridge University Press.

TRIANDIS, HARRY. 1976. Approaches Toward Minimizing Translation. *Translation: Applications and Research*, ed. by Richard. W. Brislin, 229-243. New York: Gardner Press.

WILBUR, RONNIE B. 1987. *American Sign Language: Linguistic and applied dimensions*. Boston: College-Hill Press.

WILCOX, PHYLLIS. 2000. *Metaphor in American Sign Language.* Washington D.C.: Gallaudet University Press.

What's happening with HAPPEN? The Grammaticalization of HAPPEN in American Sign Language

Benjamin Anible* and Corrine Occhino-Kehoe*[1]
University of New Mexico

ABSTRACT. In grammar books and dictionaries of American Sign Language (ASL), the word HAPPEN is generally described as a conjunction. We challenge the idea that HAPPEN functions as a 'conjunction' and instead propose an analysis of meaning and form which results in the conclusion that HAPPEN is rather functioning as an evidential marker which is grammaticalizing from the canonical verbal use. Based on the Morford, MacFarlane Corpus (2003) of 4,000 words in ASL, as well as ASL video blogs (VLOGs), interviews, and public service announcements (PSAs) collected from public YouTube channels, we extracted 50 tokens of HAPPEN from various native signers. Our analysis bears out at least three distinct uses of HAPPEN based on syntactic distribution; verbal (34%), nominal (12%) and what we will call an evidential marker (54%); denoted respectively by the following notation: HAPPEN, HAPPEN+, and HAPPEN$_1$. In addition to varied syntactic distribution, we also observe variations in phonological form and a shift in semantics toward a more subjective meaning in HAPPEN$_1$. We conclude that the canonical form of HAPPEN is undergoing a grammaticalization process evidenced by reduction in the phonological form, syntactic constriction, and semantic bleaching.

Keywords: signed languages, grammaticalization, usage-based grammar, discourse markers, conjunctions, evidentials

1. INTRODUCTION. The study of grammaticalization in signed languages is a field with potential to illustrate how diachronic changes are manifested in the lexicon of a visual modality. Though spoken language linguists have documented countless examples of grammaticalization, little work has been completed on understanding this process in signed languages. Pioneers in the field of signed language grammaticalization have shown that, like spoken languages, signed languages follow similar paths of grammaticalization with the additional caveat that signed languages also have a modality-specific pathway to grammaticalization originating in gesture rather than from within the lexicon itself (Armstrong et al. 1995, Janzen 1999, Janzen & Schaffer 2002, Wilcox et al. 2000, Wilcox 2007, Schaffer & Janzen 2012).

The set of ASL lexical items HAPPEN, UNDERSTAND, HIT, FIND, FINISH, and WRONG have traditionally been analyzed as conjunctions (Fischer & Lillo-Martin 1990, Valli 2005). Second language learners of ASL are taught that these signs occur in coordinate positions and link the content from the first clause to the second clause. The following constructed examples of HAPPEN and HIT from Humphries & Padden (1994:188) illustrate the traditional analysis for this group.

(1)   PRO.1 WALK, **WRONG** RAIN
       'As I was walking, **it suddenly** started raining.'

(2)   PRO.3 GO-AWAY VACATION, **HIT** LAID-UP SICK.
       'He went on vacation **then unexpectedly** became ill.'

(3)   PRO.1 CHAT **HAPPEN** ₃TELL₁ PRO.3 FROM WASHINGTON.
       'I was talking with him **and it happened** he told me he was from Washington.'

       While we agree with the glossing of the examples listed above, we have documented, alongside other linguists (Janzen 2007, Fischer & Lillo-Martin 1990), that the sign HAPPEN exhibits unique phonological, syntactic and pragmatic characteristics not included in Humphries & Padden's description of the sign. More specifically, the distribution of HAPPEN varies depending on the grammatical function the sign plays in a sentence. The three distinct phonological forms found in our data seem to correspond to three distinct functions: noun, verb, and evidential marker. Moreover, our data suggests that the use of HAPPEN as an evidential marker has grammaticalized from the original verbal form *to happen* and become phonologically and syntactically restricted, while also undergoing semantic modification.

2. BACKGROUND

2.1. GRAMMATICALIZATION. Using large corpora of primary source data has allowed linguists to look at language as it naturally occurs, within a discourse context. Corpus-based studies of linguistics can show how language truly functions, not just the linguist's best guess. It is through such an approach that the study of grammaticalization has arisen. Although the term arose in the early twentieth century, our current theoretical conception of grammaticalization was not formed until the end of the twentieth century by linguists such as Givón (1971), Lehmann (1983), and Traugott (1989), among others.

       Grammaticalization seeks to explain the nature of grammar through an examination of how it is created over time (Bybee 2006:712), no longer treating grammar as a synchronous set of rules. Research on grammaticalization has helped us to see grammar as a living, breathing organism, dynamic and emergent. In his 1991 article, Hopper suggests five principles which are indicative of the early and medial stages of grammaticalization: LAYERING, DIVERGENCE, SPECIALIZATION, PERSISTENCE and DE-CATEGORIALIZATION. Here we will briefly review the definitions of these principles of grammaticalization, in order to set the stage for our analysis of HAPPEN as it shows evidence of grammaticalization.[2] The following definitions are based on Hopper's original explanation for each principle.

       1) Layering: Refers to the idea that synchronically, several forms may fulfill a single
              linguistic function. New functions are added to the older functions so that
              additional layers emerge within the functional domain.
       2) Divergence: Describes the state of affairs after a split in meaning occurs.
              'When a lexical form undergoes grammaticalization to a clitic or affix, the
              original form may remain as an autonomous lexical element and undergo the
              same changes as ordinary lexical items' (22).
       3) Specialization: Is the decrease in scope of the meaning of a lexical item as it
              grammaticalizes. Eventually, specialization leads to a generalized grammatical
              meaning, different in specificity and scope from the original lexical form.
       4) Persistence: Conveys the idea that during the process of grammaticalization, it can be
              expected that a form may be polysemous, and that one or more of its meanings
              will reflect a dominant earlier meaning.

5) De-Categorialization: Describes the process that occurs as a lexical item moves toward grammaticalization, and experience a semantic shift. Generally, this can be the loss of noun-like or verb-like grammatical markers such as case or tense, an item becoming more syntactically constrained as compared to their lexical progenitor, or the loss of the autonomy afforded to members of its grammatical class, and often, the acquisition of discourse-type functions.

2.2. GRAMMATICALIZATION IN ASL. Cross-linguistically, researchers have proven that grammaticalization is a universal tendency. Looking at the process of grammaticalization in signed languages is a burgeoning field (e.g. Janzen 1995, Wilcox 2007). One of the tribulations of grammaticalization studies in signed languages is the lack of historical documentation, often going back only hundreds, rather than thousands, of years. There are precious few resources available documenting historical forms of ASL. Material comes from early illustrated dictionaries of sign language (Long 1910), archival video, and dictionaries of Old French Sign Language, as the two languages are intimately related and share many cognates.

Regardless of the limited historical resources, linguists have made progress in describing grammaticalization processes in signed languages. Wilcox (2007) has shown that signed languages have two different pathways through which grammaticalization can progress. While the first pathway is not unlike the grammaticalization process in spoken language, in which a lexical item can become a grammatical item, the second pathway seems to suggest that signed languages can skip over the lexical stage and move directly from gesture to a grammaticalized item. This is a type of grammaticalization not seen in spoken languages. An example of this is the gesture of surprise, eyes wide open, eyebrows raised, perhaps head tilted forward or mouth open. Wilcox (2007) makes the connection between this surprise gesture and the marking of a polar question. It seems that when you ask a polar question you are anticipating a response and thus the gesture of raised eyebrows or widened eyes naturally accompanies the question. This natural expression then becomes grammaticalized over time to be the grammatical marker of a polar question. Janzen (1999) has proposed that this grammatical marking for polar questions in ASL has grammaticalized over time to also become a grammatical marker for topic.

The connection between these two seemingly different constructions of the polar questions and topic marking in ASL is in the idea of shared information between the signer and the addressee. In both constructions it is clear that some amount of information is shared between the speaker and addressee concerning previously accessible and identifiable content. While there is a loss of interactivity between the signer and addressee in the topic marking construction, in that no response is required, the topic construction still functions similarly to the polar question function, in that it assumes that there is previous information that is accessible and identifiable to the addressee.

2.3. TOPICALIZATION. The way in which speakers of a given language package information and organize discourse units depends largely on how the language codifies such information. Sandler and Lillo-Martin (2006) suggest that one such way is a topic/comment construction. In some languages, such as English, topics are grammatically tied to the subject; such languages are often dubbed subject prominent. In subject-prominent languages the subject is a fully grammaticalized category and functions as the clause topic. This is not, however, a language's only option; in fact, topic-prominent languages do not have a formalized subject category but instead have grammaticalized the topic/comment construction itself.[3]

ASL has long been considered an SVO language, albeit one which makes heavy use of topicalization as an information packaging strategy. Janzen (1995) has argued that ASL has more in common with Mandarin, in terms of sentence structure of topic marking, than it does with English, and concludes it is in fact a topic-prominent language.

2.4. TENDENCIES OF TOPIC MARKING. In his 1978 treatment of conditionals and topics cross-linguistically, John Haiman convincingly demonstrated that disambiguation of the two is spurious. In actuality, topics and conditionals often share a common definition, common form (phonologically),[4] and common function (syntactically), with both identifying a frame of reference in which the main clause is either true or false. Haiman proposed that the reason these similarities are not often recognized in English is because of a superficial difference between *if* and *given that*. In reality, conditionals add to your stock of knowledge irrespective of whether the proposition is true or false, as is the case with *if*. In other words, a conditional construction selects a single instance from the sum of all possible worlds.

Strong support for the argument that HAPPEN is functioning as an evidential marker in ASL comes from Haiman's treatment of Chafe (1976), where discourse is considered a stage upon which a speaker introduces themes; 'The topic sets a spatial temporal or individual framework (...) which limits the applicability of the main prediction to a certain restricted domain' (Haiman 1978:585). We find the topic marked use of HAPPEN to function in this way.

2.5. TENDENCIES OF SEMANTIC AND PRAGMATIC CHANGE. In the following we discuss tendencies of semantic and pragmatic meaning common to grammaticalization processes and their relation to the observed functions of HAPPEN. According to Traugott 1989, propositional meanings may evolve over time to become textual (lending cohesion to a discourse) or expressive (having pragmatic or presuppositional meaning). We see this tendency for HAPPEN in our data.

Traugott represents this tendency schematically as: propositional > ((textual) > (expressive)). An example is the progression of an Old English (OE) collocation through Middle English (ME) to Present Day English (PDE), as seen in figure 4 below.

(4)   while be   while   while
        'at the time that'   'during'   'although'
        OE   ME   PDE

Traugott extends her hypothesis to propose there are highly general principles underlying lexical and grammatical items' change into expressions of epistemicity. The first part of this argument is that paths of semantic change treat epistimicity and evidentiality in identical ways. There is a degree of nebulousness in disambiguating these terms. Deontic meanings have to do with will, obligation and permission while epistemic meanings with knowledge, belief, possibilities and probabilities. Evidentiality and epistimicity both arise from auxiliary verbs, speech act verbs and modal adverbs. They both mark the state of the speakers' belief/knowledge and both mark the source of information. In summary, deontic and epistemic modals are concerned with the nature and source of knowledge; hence, despite differences in marking time and place vs. perception, they have more similarities than they do differences.

Pragmatically we tend to observe a strengthening of informativeness and relevance when lexical items undergo deontic and epistemic changes. According to Traugott (1989:51), 'Pragmatic strengthening & relevance as I use the terms largely concern strategic negotiation of

speaker-hearer interaction and, in this connection, articulation of speaker attitude.' The paths of semantic change outlined are as follows:

Tendency I: External Situation > Internal Situation (e.g. 'touch').

(5) a. John touched the jar.
     b. John was touched by the melody.

Tendency II: External/Internal Situation > Textual/Metalinguistic (e.g. 'observe').

(6) a. John observed the sails of the ships.
     b. Observe that John is a shipwright.

Tendency III: Meanings become more based in speaker subjective belief-state/attitude toward the proposition (e.g. 'while').

(7) a. John waited for a while at the harbor.
     b. While that may be true, John disagrees.

Furthermore, these tendencies proceed consecutively over the course of semantic shift. It is our analysis that HAPPEN is in the process of undergoing the third kind of change. In its grammaticalized form, HAPPEN exhibits both pragmatic strengthening—it signifies the following content as relevant and an increase in subjectivity; based in the speaker's own experience and emotions. As an evidential marker, HAPPEN also seems to function as evidential, providing information about the source of the utterance and its verifiability.

3. METHODOLOGY

3.1. DATA COLLECTION. ASL does not have a large, publicly available corpus from which to collect tokens of native speech; as a result, ASL researchers are forced to create their own corpora. One such way is to compile a collection of publicly available ASL video recordings which can later be transcribed and analyzed. We extracted 50 tokens of HAPPEN, a small number from the Morford – MacFarland Corpus (2006) of 4,000 ASL words, as well as a larger portion from Video Blogs (VLOGS), ASL interviews, and public service announcements (PSAs) collected from YouTube channels, which are open access and publically available. Speaker nativeness was judged based on conversational content including the VLOGers discussions of growing up with Deaf parents and being immersed in Deaf culture, having experiences as a Deaf person growing up, or from biographical information listed in the VLOGers online biographical sketches associated with their YouTube channel.

3.2. DATA TRANSCRIPTION. The conversations were transcribed using ELAN annotation software, originally for language and gesture studies, now widely used by signed language linguists. ELAN has grown in popularity in recent years in signed and spoken language corpus-based studies due to the many advantages it offers over traditional methods. ELAN allows real time, synchronized annotation of the primary source rather than the traditional creation of a separate transcription file, which inevitably becomes the primary reference for analysis, even though it is a secondary source.

Data were tagged for grammatical function, phonological form, and syntactic position as well as any additional pragmatic information added to the usage event for each token of HAPPEN.

4. RESULTS. From the 50 transcribed tokens, three distinct uses of HAPPEN arose; namely verbal, nominal and an evidential marker; each of which had distinct phonological forms. Each type of HAPPEN is denoted by the following notation: HAPPEN for the canonical verbal form, HAPPEN+ to denote the nominal form (often glossed as EVENTS) and HAPPEN$_1$. Description and distribution shown in Table 1:

Gloss	HAPPEN	HAPPEN+	HAPPEN$_1$
Tokens (n = 50)	17	6	27
Part of speech	Verb	Noun	Discourse marker
Translation	to happen	events	then, one time, once
Syntactic dist.	Unrestricted	Unrestricted	Clause initial
Phonological marking	[pivot] [orientation]: *pronation → abduction*	[pivot] [orientation]: *pronation → abduction* [repeat]	*either/or* [pivot] only [orientation] only: *pronation → abduction*

TABLE 1. Description and Distribution of HAPPEN data.

Examples of their use in context are as follows:

(8)   PRO.3>wife SUPPOSE SOMETHING **HAPPEN** TO PRO.3
      (wife saying): 'What if something had happened to you?' (Letsgofly08)

(9)   PRO.1 SIGN TELL-STORY THAT FUNNY **HAPPEN**+
      I was telling a story about some funny events.' (KeithWann)

(10) **HAPPEN₁** PRO.1 WORK FOR GALLAUDET (...)
      'I was working for Gallaudet University.' (Dndmartin)

4.1. PHONOLOGICAL REDUCTION. Our data shows a progression of reduction from HAPPEN to HAPPEN$_1$ in which the verb is rarely reduced but evidential marker forms are frequently reduced. The forms HAPPEN and HAPPEN+ are examples of a common process in ASL identified by Supalla & Newport (1978) as VERB/NOUN PAIRS. In these pairs, semantically related processes and things (verbs and nouns) are differentiated in their form by a larger single movement for verbs and smaller reduplicated movements for nouns. Examples of this verb/noun paradigm from ASL are PRINT/NEWSPAPER, CHIRP/BIRD, SIT/CHAIR, and FLY/AIRPLANE. It is common in these pairs for the reduplicated movement to be partially reduced in manner or

degree. For example in SIT/CHAIR, the pointer and middle finger are bent on the dominant hand (representative of bent legs) which then move downward to contact the same two fingers of the non-dominant hand (representative of a chair seat). In the verbal form the distance the dominant hand travels before contact is greater than any single token of the noun's reduplicated movement. Additionally, we can describe the verbal movement as more PROXIMALIZED while the nominal movements are more DISTALIZED.

We see this same tendency in HAPPEN/HAPPEN+. The degree of movement in the noun is reduced as compared to that of the verb. In our data we found variation where the full form has both a twisting articulation at the elbow and a twisting articulation at the wrist. This form is based on the earliest documented uses of verbal HAPPEN (Long 1910). This twisting often reduces to wrist only articulation in the nominal and evidential marker tokens of HAPPEN. The parameters that vary in our analysis are PRONATION, when the hand and wrist have not been modified (i.e. are at a neutral position with wrists facing inward as they would be held at rest), and ABDUCTION, in which the hand is facing down without movement of the wrist, but rather with a movement of the forearm which originates from articulation at the elbow.

FIGURE 1. Early attested form of HAPPEN (Long 1910).

The manner of reduction seen in HAPPEN+ is often identical to that in HAPPEN$_1$ with the notable exception that HAPPEN$_1$ has no reduplication of movement. Either the abduction or pronation may be elided. One type of phonological reduction, common across signed languages occurs when the movement of a sign, which is canonically produced closer to the body (proximally), is produced further from the body (distally) (Brentari, 1998). Thus, we should expect the absence of abduction to be the more common form of reduction (as opposed to pronation) as well as an indication of a higher degree of reduction.

We found no clear tendency in the direction of either abduction or pronation absence and so categorized either type simply as phonological reduction. We did find significant reduction in phonological form for evidential markers 63% (17/27) compared to 6% (1/17) for verbs and 0% 0/6 for nouns (p < .001).

	HAPPEN/HAPPEN+	HAPPEN$_1$
Phonological full form	22	1
Phonological reduction	10	17

TABLE 2. The relation between phonological reduction verses discourse uses of HAPPEN.

4.2. SYNTACTIC RESTRICTION. Neither HAPPEN nor HAPPEN+ show restrictions in their possible syntactic environments. We found both verbal and nominal forms of the sign in phrase medial and final positions. Our analysis does not preclude them from occurring phrase initially but in our data we did not encounter any such instances. In the following we see HAPPEN used medially in example 11 and finally 12 as well as HAPPEN+ used medially 13 and finally 14 while the evidential marker HAPPEN$_1$ overwhelmingly occurs phrase initially.

(11) PRO.1 WANT INFORM ONE STORY **HAPPEN** TRUE-BIZ HAPPEN
TERRIBLE EXPERIENCE
'I want to tell you about one true story that **happened** to me, was a
terrible experience.' (Dndmartin)

(12) #WHAT **HAPPEN**?
'What **happened**?' (Morford-MacFarlane)

(13) THAT TRUE CONNECTION U-N-I-T-Y THAT FINISH **HAPPEN+**
UP-TIL-NOW FINISH SEE+
'Now that is true unity, **those things** you often see.' (Ella Mae Lentz)

(14) PRO.1 SIGN TELL-STORY THAT FUNNY **HAPPEN+**
'I was telling a story about some funny **events**.' (Keith Wann)

Any exceptions to the initial position of HAPPEN$_1$ were instances in which the preceding constituent is a time adverbial. Pragmatically, time adverbials and HAPPEN$_1$ perform similar functions, so it is unsurprising to find their order to be interchangeable as in 15. Additionally, HAPPEN$_1$ was followed by an NP in most cases (n=21). In the few instances where we did not find HAPPEN$_1$ to be followed by an NP (n=24), time adverbials present. Such adverbials appear to be interchangeable with HAPPEN$_1$ as they serve very similar functions. Alternately, we did have an additional three instances where the following consitutent was a verbal phrase (VP). We were forced to exclude these tokens due to low token frequency; there were not enough instances to form any kind of generalization. We believe that HAPPEN may have a fourth grammatical use, that is, when used in these pre-verbal phrases it seems to be functioning as a type of modal or auxiliary verb. The approximate translation in such instances may be 'was' or 'did' HAPPEN. Though we have too few tokens to make strong predictions about this other usage, modals and auxiliaries are common candidates for grammaticalization processes, and including this use in our grammaticalization argument would not be counter-intuitive (see 15 for an example).

(15) ALWAYS **HAPPEN₁** HIT ONE HEARING MAN STARE MEAN MOCK DEAF
      'Then it always **happens that** suddenly a mean hearing man will see the fire and mock
      the Deaf.' (Olsenterp)

Here the ASL sign HIT is not the verb 'to hit', but the adverb HIT which commonly is glossed as SUDDENLY. Here our token of HAPPEN occurs between two adverbs, suggesting an additionally restricted syntactic environment for the reduced form.

4.3. SEMANTIC/PRAGMATIC SPECIALIZATION. We find HAPPEN₁ to function as an evidential. Following Traugott's (1989) definition of evidentials, HAPPEN₁ serves to mark the source of information and the speaker's belief about the validity of that source. Namely HAPPEN₁, occurring before the recitation of a narrative, tells the interlocutor that the speaker has a personal interaction with the content following and that they believe it to be true. HAPPEN₁ also clearly grounds the statement in the past (except in conditional use).

Additionally, subjectivity is common of evidentials. We suspect HAPPEN₁ along with the other 'conjunctions' discussed in the introduction of this paper, give information about the emotional state of the speaker. This tendency is clear with examples like WRONG or HIT where distress and surprise are conveyed respectively. With HAPPEN₁ subjective content is slightly more nebulous, but there are intimations of speaker interest in the topic, and a desire to convey relevance. The marking of relevance is one of the properties of Traugott's third tendency of semantic change.

Using HAPPEN₁ at the beginning of a story or phrase alerts the listener as to what the speaker wishes to talk about. This is clearly the domain of topic marking; therefore, unsurprisingly we find near total co-occurrence of HAPPEN₁ and topic marking in our data. Janzen (2007) has called these FORWARD-LOOKING or TRANSITION PIVOTS that suggests they add coherence and cohesion to the discourse. With the following chain of relations in mind we conclude that HAPPEN₁ is a grammaticalized evidential marker. HAPPEN₁ exhibits reduction in form, reduction of syntactic distribution, specialization of its semantic content, and functions in discourse significant constructions.

This type of grammaticalization process is not limited to ASL. The verb, 'to happen,' exhibits a crosslinguistic tendency toward grammaticalization into an evidential marker, or marker of evidentiality. We see evidence of a similar construction grammaticalizing in Curnow and Travis (2009:22). The Spanish verb *pasar* (to happen) appears in a similar context to HAPPEN₁ as a focus element in pseudo clefts.

(16) *lo que pasa es que estaba= mal conectado*
      'The thing (lit. 'what happens') is that it is badly connected.' (Curnow and Travis 2009)

Significant pragmatic strengthening, following the definition discussed above, was found for the reduced forms of HAPPEN more often than for unreduced forms; 100% (27/27) for HAPPEN₁, compared to 6% (1/17) for verbs (HAPPEN) and 0% (0/6) for nouns (HAPPEN+) (p < .001). The associations between the forms (reduced vs. full) and semantic-pragmatic use (verb/noun vs. discourse uses) is illustrated in Figure 2. The size of the circles corresponds to the frequency of the given type. The thickness of lines indicates the strength of the association, based on the numbers in Table 2. The strongest association are from reduced forms to discourses uses (HAPPEN₁)and full forms to canonical uses (HAPPEN and HAPPEN+), supporting our

categorization of form and meaning relations (Table 1). In this way variable discourse situations can be resolved toward disambiguition where an innovative phonetic form has a narrowed meaning/sense. This is precisely the traditional characterization of grammaticialization processes (Bybee 2006:726).
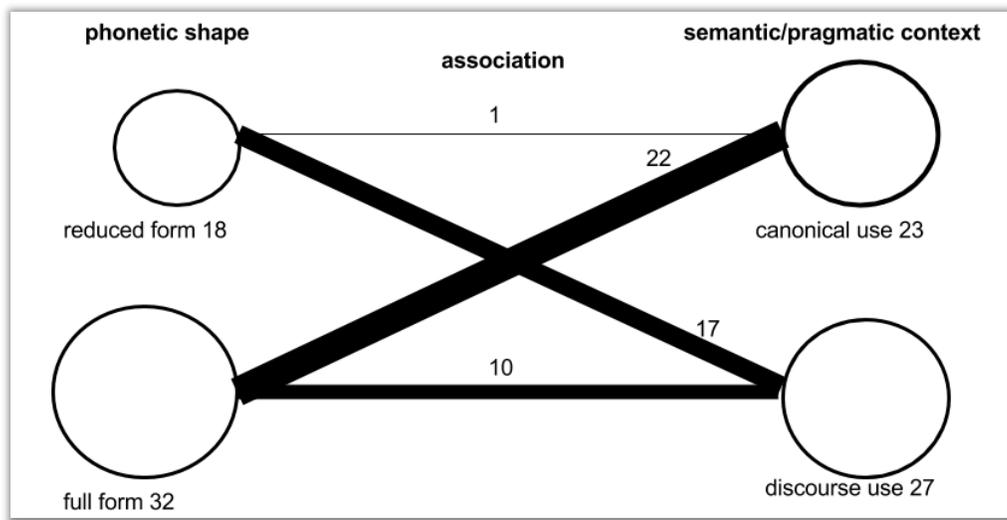
FIGURE 2. Phonetic and semantic/pragmatics in tokens of HAPPEN.

5. CONCLUSION. Though our sample size is smaller than most literature on grammaticalization processes we found a clear tendency for HAPPEN$_1$ to function as an evidential marker, possibly including co-occurrence of topicalization. In the future, as larger ASL corpora become available, we would like to test this tendency against a bigger corpus to see if the pattern persists. Janzen (1999) and Fischer & Lillo-Martin (1990) have demonstrated the tendency of other verbs to follow a similar path as HAPPEN, namely FINISH and UNDERSTAND respectively, shifting their function and usage toward a discourse level marker. All three of these verbs belong to the same category; signs previously considered to be conjunctives. While this usage as a conjunctive may have been the primary stage of their grammaticalization, from verb to subordinating conjunction, it seems clear that these words have continued to grammaticalize, picking up more pragmatic function along their path. Fischer & Lillo Martin (1990) have suggested that other verbs, such as WRONG and SUPPOSE, have also made the transition to subordinating conjunctions. If we are to assume that the shift toward subordinating conjunction is only the first step in the process of grammatical change, perhaps these word too have followed FINISH and HAPPEN to become further grammaticalized into evidential marker. Regardless, it is clear that the grammaticalization path from full content word to functional item – well established cross-linguistically in the spoken modality – holds true within the manual modality of ASL as well.

NOTATIONAL CONVENTIONS: IDENTIFYING SIGNS

Symbol	*Example*	*Use*
SIGN	CAT	A single uppercase English word identifies a single ASL sign. Using the gloss CAT to identify a sign 'cat' does not mean that the sign has the same morphological, syntactic, or semantic characteristics as the English word *cat*.
SIGN-SIGN-SIGN	OH-I-SEE	Uppercase English words separated by hyphens also represent a single sign.
SIGN^SIGN	HOME^WORK	The symbol ^ indicates that two signs have been combined into a single compound sign.
SIGN+++	TREE++	Plural form of a noun, denoted by repeated morphemic movement.
SIGN+	HAPPEN+	
S-I-G-N	B-U-D-G-E-T	Hyphens between uppercase letters indicate a sequence of alphabetic character signs used to spell a word.
#SIGN	#WHAT	A word beginning with the symbol # indicates a lexical sign whose origin is ultimately traceable to a sequence of alphabetic character signs.
PRO.#	PRO.1	The notation -1 indicates a first person form. Number represents person.
POSS.#	POSS.1	Denotes a possessive form. Number represents person.
SIGN$^{[ASPECT]}$	WORK$^{[DURATIONAL]}$	A sign derived from WAIT. The label in the square brackets identifies the grammatical process underlying the sign being represented.
SIGN$_1$ . . . SIGN$_n$ THEME----------	EVERY-WED CALENDER------	The top line represents the signing of the strong hand while the bottom line represents the signing of the weak hand. Here the weak hand maintains the THEME buoy in place during the sign sequence SIGN$_1$ . . . SIGN$_n$ This serves as a form of anaphoric schematization.

'gesture'	'wave'	Denotes a gesture without specified lexical form.
(a) _____ x SIGN$_1$ . . . SIGN$_n$ (b) [SIGN$_1$ . . .SIGN$_n$]-x	(a)_____ t BOOK RED (b) [BOOK RED]-t	The sequence of signs from SIGN$_1$ to SIGN$_n$ is accompanied by a non-manual signal, in this case for topicalization. Other non- manual signal categories include n (negation), q (y/n, rhetorical, wh question), cond (conditional), and certain adverbials. [1]

NOTATIONAL CONVENTIONS: DIRECTION AND PLACEMENT

Symbol	*Example*	*Use*
(a) PRO$_{\rightarrow x}$ (b) [PRO]>x	(a) PRO$_{\rightarrow girl}$ (b) [PRO]>girl	$\rightarrow$x indicates that the sign is directed toward entity x. Entity x will be either an element of real space or a real-space blend. [1]
PRO$_{\rightarrow |x|}$	PRO$_{\rightarrow |aunt|}$	$\rightarrow$|aunt| indicates that the sign is directed toward the blended entity |aunt|.
PRO-PL$_{a - b - c}$	THEM$_{John - Mary - Ed}$	a - b - c indicates that the hand moves along a path, such that the extent of the path points toward entities a, b, and c.
$_x$SIGN$_y$	$_3$GIVE$_1$	Indicates that the sign begins nearer to and/or directed toward x, then moves toward y.
[SIGN]$_{\rightarrow L}$	[COLLEGE]$_{\rightarrow L}$	The nondirectional sign enclosed the square brackets is directed toward location L.

REFERENCES

ARMSTRONG, DAVID F.; WILLIAM C. STOKOE; and SHERMAN WILCOX . 1995. *Gesture and the nature of language*. Cambridge: Cambridge University Press.

BRENTARI, DIANE. 1998. *A Prosodic Model of Sign Language Phonology*. Cambridge, MA: MIT.

BYBEE, JOAN. 2006. From usage to grammar: The mind's response to repetition. *Language* 82.711–733.

CHAFE, WALLACE. 1976. Givenness, contrastiveness, definiteness, subjects, and topics. *Subject and Topic,* ed. by Charles N. Li, 25–56. New York: Academic Press.

CURNOW, TIMOTHY JOWAN, and CATHERINE E. TRAVIS. 2009. Clause-combining in Spanish cleft constructions. Paper presented at the 11th Conference of the International Pragmatics Association (IPrA), Melbourne University, Melbourne.

FISCHER, SUSAN, and DIANE LILLO-MARTIN. 1990. Understanding conjunctions. *International Journal of Sign Linguistics* 1.71–80.

GIVÓN, THOMAS. 1971. Historical syntax and synchronic morphology: An archaelogist's field trip. *Chicago Linguistic Society* 7.394–415.

HAIMAN, JOHN. 1978. Conditionals are topics. *Language* 54.564–589.

HOPPER, PAUL. 1991. On some principles of grammaticalization. *Approaches to Grammaticalization*, ed. by Elizabeth Closs Traugott and B. Heine, 1.17–35. Amsterdam: John Benjamins.

HUMPHRIES, TOM and CAROL PADDEN. 1994. *A basic course in American Sign Language* (2nd ed.). Silver Spring, Maryland: T.J. Publishers.

JANZEN, TERRY. 1995. *The polygrammaticalization of FINISH in ASL*. Manitoba, Ontario: University of Manitoba master's thesis.

JANZEN, TERRY. 1999. The grammaticization of topics in American Sign Language. *Studies in Language* 23.271–306.

JANZEN, TERRY. 2007. The expression of grammatical categories in signed languages. *Verbal and Signed Languages: Comparing Structures, Constructs and Methodologies* (vol. 36), ed. by Elena Pizzuto, Paola Pietrandrea, and Raffaele Simone, 171–197. Berlin: Mouton de Gruyter.

JANZEN, TERRY and BARBARA SHAFFER. 2002. Gesture as the substrate in the process of ASL grammaticization. *Modality and structure in signed and spoken languages* 199–223.

LONG, JOSEPH SCHUYLER. 1910. *The sign language: A manual of signs*. Washington, D.C.: Press of Gibson Bros.

LEHMANN, CHRISTIAN. 1983. Universal and typological aspects of agreement. *Apprehension: Das sprachliche Erfassen von Gegenstiinden* (vol. 2), ed. by H. Seiler, & Franz J. Stachowiak, 201–267. Tübingen: G. Narr.

MORFORD, JILL, and JAMES MACFARLANE. 2003. Frequency Characteristics of American Sign Language. *Sign Language Studies* 3.213–225.

SANDLER, WENDY, and DIANE LILLO-MARTIN. 2006. *Sign language and linguistic* universals. Cambridge UK; New York: Cambridge University Press.

SHAFFER, BARBARA and TERRY JANZEN. (2012). Gesture, Lexical Words, and Grammar: Grammaticalization Processes in ASL. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* (vol. 26), University of California, Berkeley.

SUPALLA, TED, and ELISA NEWPORT. How many seats in a chair? The derivation of nouns and verbs in American Sign Language. *Understanding language through sign language research*, ed. by Patricia Siple, 91–132. New York: Academic Press.

TRAUGOTT, ELIZABETH. 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 65.31–55.

VALLI, CLAYTON. 2005. *Linguistics of American Sign Language: An introduction*, *4th ed*. Washington D.C.: Clerc Books.

WILCOX, SHERMAN; BARBARA SHAFFER; MARIA J. JARQUE; J. VALENTI; ELENA PIZZUTO; & PAOLO ROSSINI. (2000). The emergence of grammar from word and gesture: A cross-linguistic study of modal verbs in three signed languages. 7th International Conference on Theoretical Issues in Sign Language Research, Amsterdam.

WILCOX, SHERMAN 2007. Routes from gesture to language. *Verbal and Signed Languages: Comparing Structures, Constructs and Methodologies*, ed. by Elena Pizzuto, Paola Pietrandrea, and Raffaele Simone, 107–131. Berlin: Mouton de Gruyter.

PRIMARY DATA SOURCES

KEITH WANN. 1 of 4 interview with Keith Wann on the learning and laughter with Louise Radio Show. *YouTube*. Web. 15 Nov. 2010.

OLSENTERP. Deaf camping. *YouTube*. Web. 15 Nov. 2010.

OLSENTERP. Deaf church. *YouTube*. Web. 15 Nov. 2010.

ELLA MAE LENTZ. Deaf community divided. Whose fault is it? www.ellasflashlight.com. Web. 15 Nov. 2010.

DNDMARTIN. Deaf person and cop story. *YouTube*. Web. 31 Oct. 2010.

LETSGOFLY08. TRUE STORY about two deaf couple on 9/11. *Youtube*. Web. 31 Oct. 2010.

LANCEPICKETT. VP3D: The videophone of the future? *YouTube*. Web. 15 Nov. 2010.

WANN, KEITH. Keith Wann interview. *Here and Now*. NPR.

NOTES

[1] Authors listed alphabetically.
[2] For a more in depth discussion of these 5 principles of grammaticalization see Hopper 1991).
[3] Janzen (1999) provides a description of Li and Thompson's article on formalized topic-comment structure in Mandarin Chinese and its relation to ASL topic/comment structure.
[4] Of course this varies typologically, though topic and comment do have a strong tendency to be marked the same within a given language.

Describing reduplication patterns in Tohono O'odham with language learners in mind

Keiko Beers, Robert Cruz, Laura Hirrel, Iphigenia Kerfoot
University of New Mexico

ABSTRACT. Tohono O'odham (TO) has a complex reduplication system expressing number and aspect. However, there is no comprehensive explanation of TO reduplication patterns in the literature that is consistent with our data. Additionally, the only pedagogical grammar available for the language presents a single reduplication pattern, although several patterns are productive. Our study examines common reduplication processes, proposing that they conform to a limited and predictable set. We suggest that the prosodic weight of the reduplicant is dependent upon the weight of the base. Furthermore, we suggest that this generalization is useful for language learners to understand because it allows them access to a larger number of reduplication patterns in their endangered heritage language.

Keywords: Tohono O'odham, O'odham, reduplication, prosodic morphology, L2

1. INTRODUCTION. Tohono O'odham (hereafter, TO), an endangered Uto-Aztecan language spoken primarily in southern Arizona and northern Sonora, Mexico, has a complex reduplication system used for the expression of grammatical number (both plural and distributive) as well as the iterative aspect (Fitzgerald 2003). Although reduplication is a productive and highly salient morphological process in TO, the literature does not provide a comprehensive description or explanation of the language's reduplication patterns that is consistent with our data (e.g. Hill & Zepeda 1992; Fitzgerald 1999, 2003; Riggle 2006). Furthermore, the only pedagogical grammar available for the language (Zepeda 1983) exclusively describes a single reduplication process, which is only able to account for a subset of the reduplication patterns used in the language. Given the endangered status of TO, a presentation of common reduplication patterns would be beneficial to second language (L2) learners. At first glance, however, reduplication in TO may appear random and unpredictable to these learners. The aim of the present work is to present a unified framework that is able to account for more patterns of reduplication in TO than previously covered by the literature. Our account provides a straightforward explanation of many of the reduplication patterns and can be easily implemented in the classroom.

This study examines and describes the primary processes evident in the data we have collected for TO. We propose that reduplication conforms to a limited set of predictable patterns, and we rely on prosodic theory to provide a unified understanding of TO reduplication. Our analysis contrasts with the account given by Zepeda (1983:7) in which reduplication is described as 'CV-copying'. For example, ***goks***[1] 'dog' → ***gogoks*** 'dogs' would be explained as the copying of the initial CV (i.e. **C**onsonant, **V**owel) in the singular form of the word (i.e. *go*). While the pattern of CV-copying was likely used by Zepeda for pedagogical purposes, this description does not account for other reduplication patterns in the language. There are other cases in which the reduplicant (the portion of the word that is reduplicated in the plural) either lengthens (***ban*** 'coyote' → ***ba:ban*** 'coyotes') or shortens (***mi:sto ɺ*** 'cat' → ***mimisto ɺ*** 'cats'). This study suggests that the length of the reduplicant can be understood in terms of a prosodic pattern, termed 'quantitative complementarity' (cf. §2.1) by McCarthy and Prince (1995:334–5). When this theory is considered within the context of TO, the reduplicant occurs in a prosodically light (monomoraic) form when paired with a heavy (polymoraic) base, and in a heavy form when

paired with a light base. The length of the reduplicant in turn causes structural changes in the base (e.g. compensatory lengthening, vowel reduction, syncope, etc. (cf. §§2.2 and 2.3)).

Our primary tasks in this paper are two-fold. First, we illustrate how McCarthy and Prince's (1995) theory of prosodic morphology accounts for many of the reduplication patterns in our data (§§2.2 and 2.3). Secondly, we briefly provide suggestions for how our theoretically based findings can be practically applied in the classroom (§3). In closing, we propose some avenues for future research (§4).

1.1. LANGUAGE BACKGROUND. The TO language belongs to the Uto-Aztecan family, whose languages are spoken from Oregon to Montana in the north, down into El Salvador. TO is a member of the Tepiman branch along with three indigenous languages spoken in Mexico: O'ob No'ok, Ódami, and O'dam.

TO is spoken on both sides of the US-Mexico border in the states of Arizona and Sonora. In the United States this variety of O'odham has previously been called Papago. TO, along with every other indigenous language of North America, is endangered, as intergenerational transmission has been disrupted and the majority of speakers are older than forty (Fitzgerald 2010 citing Madsen 2004).

1.2. METHODOLOGY. The data in this paper are from Mr. Robert Cruz's Ko:kolo:di dialect, spoken in the Cukuḍ Kuk district (also known as Tecolote) in the southernmost region of Arizona, along the national border between the United States and Mexico. These data are drawn from a corpus which has been developed and recorded over the course of two years, supplemented by targeted elicitation. In many cases, previous works on TO (Fitzgerald 1999, 2003, 2010; Hill and Zepeda 1998; Saxton & Saxton 1969; Saxton et al. 1983; Zepeda 1983) provided ideas for elicitation, as well as consultation with Mr. Cruz.

2. DATA PRESENTATION: VOWEL PROCESSES IN TO REDUPLICATION. As stated, TO utilizes the reduplicative process to express plural nouns (***mi:stoɻ*** 'cat' → ***mimistoɻ*** 'cats') and plural verbs. In the latter case, the reduplicative pattern reflects a plural subject (***ko:ş*** 'sleeping.SG.SUB'[2] → ***ko:kş*** 'sleeping.PL.SUB'), plural objects, or an iterative action (***hi:nak*** 'barking.PUNCT' → ***hihinak*** 'barking.ITER'). For the discussion at hand, our data are solely comprised of nouns. Fitzgerald (2003:65) has also found a geminate formation in the reduplicated, derived (i.e. plural) word to express the distributive (*ta:tam*.SG 'tooth.' → *ta:ttam* 'teeth.DIST'). However, this formation is not present in Mr. Cruz's speech.

Two primary reduplicative processes are present in TO. The first is LENGTHENING OF THE VOWEL in the reduplicant (the reduplicated portion of the word). In these cases, the initial CV in the singular base lengthens to CVV in the reduplicant of the plural form (***ban*** 'coyote' → ***ba:ban*** 'coyotes'). The second pattern is SHORTENING OF A LONG VOWEL appearing in the initial syllable of the singular base. That is, a CVV becomes CV, as in ***má:kai*** 'medicine man' → ***mámakai*** 'medicine men').

There are also cases where the CV reduplicant remains unchanged from the singular to the plural as in ***dʒúḍum*** 'bear' → ***dʒúdʒəḍum*** 'bears' and (***hudʒit*** 'lizard' → ***huhudʒit*** 'lizards'). While this appears to be the most straightforward pattern, it is not the default in our data. This issue will be revisited in §2.3.

In the context of both lengthening and shortening processes, we've observed other alterations in the reduplicated forms, which will be further discussed in §2.2 and §2.3.

We propose that McCarthy and Prince's (1995) PROSODIC MORPHOLOGY HYPOTHESIS provides a possible explanation for many of the patterns that we have observed in our data (cf. §2.1). Detailed analyses and discussion of our data are given in §§2.2 and 2.3.

2.1. MCCARTHY AND PRINCE'S PROSODIC MORPHOLOGY HYPOTHESIS. McCarthy and Prince's (1995) theory of prosodic morphology provides an explanation for the various lengthening and shortening processes in TO reduplication. The PROSODIC MORPHOLOGY HYPOTHESIS suggests that PROSODIC UNITS—moras, syllables, feet, etc.—comprise the shape of 'templatic' morphemes, reduplicated segments in this case, in relation to the prosodic makeup of the base (McCarthy & Prince 1995:318–319). Templates 'must respect the well-formedness requirements of prosody' (McCarthy & Prince 1995:318). While we take the view that speakers of languages and language usage determine linguistic forms, McCarthy and Prince nonetheless offer a useful method for describing the patterns noted in TO reduplication. Their view is a departure from the assumption that specific constituents are copied in languages, for example, 'copy CV or CVC'. In regards to reduplication, our data suggest that TO functions in terms of a QUANTITATIVE COMPLEMENTARITY in which light reduplicants generally appear with heavy bases and vice versa (McCarthy & Prince 1995:334).

Before setting out to demonstrate how TO adheres to this system, an explanation of how heavy and light syllables are defined in TO is necessary. In languages such as TO, moras are prosodic units, which contribute weight to a given a syllable. Traditionally in studies of prosody, a CV segment is always treated as a light, monomoraic syllable, whereas CVV and CVC are regarded as heavy syllables with bimoraic feet. In this case, the coda consonant carries moraic weight. TO does not, however, follow this pattern. According to Hayes (1989:255), in languages in which a distinction between CV and CVV syllables exists, CVC syllables pattern as light rather than heavy (also see Tranel 1991 for a related discussion). In other words, the final consonant of the CVC structure does not contribute moraic weight. For example, *ban* 'coyote', which has a CVC structure is light rather than heavy in TO. Such languages 'differ in their rules for assigning moraic structure' (Hayes 1989:255). This explanation accounts for many of the reduplicative patterns seen in our data. The following sections aim to demonstrate how the prosodic morphology hypothesis applies to reduplication in TO.

2.2. LENGTHENING PROCESSES IN TO REDUPLICATION. Long vowel reduplication refers to the process of lengthening the vowel in the reduplicant when the initial vowel in the singular base is short. Within the context of reduplication, we hypothesize that reduplicants with long vowels can be understood as being quantitatively complementary, as suggested in McCarthy and Prince (1995) (cf. §2.1), to singular bases that are light in O'odham.[3] This current section describes vowel lengthening reduplication patterns in our data.

MONOSYLLABIC LIGHT BASES. In the data collected from Mr. Cruz's speech, reduplicants with long vowels (CVV) only occur when the singular base form is light (CVC). This pattern can be described in terms of McCarthy and Prince's (1995) principle of quantitative complementarity, where light reduplicants occur with heavy bases and heavy reduplicants with light bases. In the cases presented in Table 1, all of the singular base forms exhibit monosyllabic CVC structures and the vowel in the reduplicant is long. The vowel lengthening in the reduplicant demonstrates that the CVC base syllable patterns as light (e.g. ***ban*** → 'coyote' ****baban*** 'coyotes'), and therefore occurs with a complementary heavy reduplicant. In the cases of vowel lengthening in

the reduplicant where the singular base is monosyllabic, the resulting form is a heavy-light syllable, which is argued to be the most common pattern in TO (Hill & Zepeda 1998, Fitzgerald 2003).

Base (singular)	Reduplicated (plural)	Gloss
bán	**bá:**ban	coyote
mád	**má:**mad	child (of woman)
ṭá̰d	**ṭá̰:**ṭad	foot
ɲɨ́m	**ɲɨ́:**ɲɨm	liver
dʒɨ́g	**dʒɨ́:**dʒɨg	hole

TABLE 1. Vowel lengthening in the reduplicants.

LESS FREQUENT LENGTHENING PROCESSES. Two additional lengthening patterns are found in the data. These patterns are less frequent in our data set and likely more restricted in distribution. Nevertheless, the data in this section still exemplify patterns relevant to quantitative complementarity as described in the prosodic morphology hypothesis (McCarthy & Prince 1995). Due to the complexity of these two patterns, they will not be discussed in the APPLICATIONS IN THE L2 CLASSROOM section (§3) of this paper.

Base (singular)	Reduplicated (plural)	Gloss
nówĭ	**nó:**nhoĭ	hand
báhĭ	**bá:**bhaĭ	tail
wúhĭ	**wú:**puĭ[4]	eye

TABLE 2. Lengthening with extra-short vowels and metathesis.

In Table 2, the initial vowel in the singular base is short, while the vowel in the reduplicant form is long. The lengthening in the reduplicant in these examples suggests that the extra-short vowel word-finally, denoted by the breve, does not contribute weight to the singular base (see Hill & Zepeda 1998:357-358 on extra-short vowels). Again, the vowel in the reduplicant lengthens, complementing the light CVC base. Hill and Zepeda (1992:374) recognize the second observable phenomenon shown in the data included in Table 2 as metathesis. For example, the plural form of *bahĭ* 'tail' might be expected to be **ba:bahĭ*, however, the form appears to be affected by a process of metathesis and appears as *ba:**bha**ĭ*. The process of metathesis prevents the occurrence of laryngeal codas, a violation of TO phonotactics.[5] In some cases, as in the plural forms for 'hand' and 'tail', the process of metathesis leads to a resyllabification of the initial onset consonant in the singular base to act as a coda in the reduplicant of the plural form.

The other reduplicant lengthening pattern that occurred less frequently in the data collected from Mr. Cruz is illustrated in Table 3. At first glance, these data do not appear to follow the pattern of quantitative complementarity for reduplication. The singular base forms should be treated as heavy because they are assigned at least two moras, yet the vowel in the reduplicant exhibits vowel lengthening. One important difference between the data shown in Table 3 and those in Tables 1 and 2 is that all data tokens in Table 3 show processes of reduction and syncope (i.e. segment deletion) in the plural, reduplicated forms. For instance, in the reduplicated form of *dóʔag* 'mountain' both the first vowel occurring in the base and the

following glottal stop are absent in the plural form (*dodóʔag).

Base (singular)	Reduplicated (plural)	Gloss
dóʔag	**dó:**dag	mountain
báʔamad	**bá:**bəmad	daughter's child
móʔo	**mó:**mĭ̥	head
dʒíʔĭ	**dʒí:**dʒ	mother

<div align="center">TABLE 3. Compensatory vowel lengthening.</div>

We hypothesize that the lengthening of the vowel in the reduplicant is compensatory lengthening as a result of the loss of segments (cf. Hayes 1989). In cases like those shown in Table 3, historically, the reduplicants may have been light because the singular bases are heavy. We hypothesize that as processes of reduction in the plural bases occurred, the vowel in the reduplicant lengthened, adding more prominence to the stressed syllable. In future research, we wish to investigate the diachronic sources of the data type shown in Table 3.

2.3. REDUCTION PROCESSES IN TO REDUPLICATION. The previous section demonstrated how vowel lengthening in the reduplicant occurs in circumstances in which the base is light (CVC). This section will show how the reduplicant patterns when the base is heavy (i.e. greater than CVC), and how such patterns can be understood in terms of quantitative complementarity as described by McCarthy and Prince (1995). We will also briefly address some additional reduction changes which affect the plural forms of the bases, which are likely related to common processes of language change found cross-linguistically.

POLYSYLLABIC HEAVY BASES. The reduction process most relevant to McCarthy and Prince's (1995) theory of prosodic morphology affects the structure of the reduplicant as was the case in the data sets described in §2.2. However, in these instances, if the singular form of the word has a long vowel in the initial syllable (CVV) and the base is heavy (e.g. CVCV, CVVCV, etc), then it will frequently shorten to CV in the reduplicant as show in Table 4. Of less significance to the theory of prosodic morphology, but a pervasive pattern nonetheless, the data show that the same long vowel shortens in the plural base (e.g. *tótobĭ*, not **tóto:bĭ*) due to the phonotactics of TO, which rarely allow long vowels word internally.

Base (singular)	Reduplicated (plural)	Gloss
tó:bĭ	**tó**tobĭ	rabbit/cottontail
sí:kĭ	**sí**sikĭ	white tail deer
má:kai	**má**makai	medicine man
mí:stoɻ	**mí**mistoɻ	cat
ɻí:wa	**ɻí**ɻiwa	jacket (Sp)[6]

<div align="center">TABLE 4. Vowel shortening in the reduplicant.</div>

The outcome of vowel shortening is a light syllable occurring with a heavy polysyllabic base, as noted in other languages that demonstrate quantitatively complementary reduplication (McCarthy & Prince 1995:334).

In cases in which the vowel in the initial syllable of the singular base is already short, no

change ensues in the reduplicant (e.g. *ʧímit* 'tortilla' → *ʧíʧəmit* 'tortillas' and *hodaǐ* 'rock/stone' → *hohodaǐ* 'rocks/stones'). While this is essentially CV-copying, the most straightforward reduplication pattern, the motivation behind this in terms of quantitative complementarity is that a change does not take place because the initial syllables are already light (CV), maintaining a complementary prosodic relationship with the heavy bases (see Tables 5 and 8 below for more examples).

REDUCTION PROCESSES AFFECTING THE PLURAL BASE. We have noted a couple of additional reduction processes that affect the plural forms of the bases in the context of words that are prosodically heavy. While noting these patterns is not necessary for pedagogical purposes, they pose some interesting suggestions regarding how TO may be evolving. Furthermore, despite these alterations in the plural bases, the data are still consistent with quantitatively complementary reduplication.

In the first reduction process, the initial syllable of the base centralizes to a schwa (bolded in the data set) in the plural forms of the bases as exemplified in Tables 5 and 6. In terms of complementarity, Table 5 shows reduplicants that are identical to the initial CV forms of the singular bases, while Table 6 includes reduplicants that shorten. Both patterns transpire in order to maintain a complementary relationship of light reduplicants with heavy bases.

Base (singular)	Reduplicated (plural)	Gloss
ʤúdʉm	ʤúʤədʉm	bear
ʧímit	ʧíʧəmit	tortilla
náwaṣ	nánəwaṣ	pocket knife (Sp)
kámiṣ	kákəmiṣ	shirt (Sp)

TABLE 5. Reduction of full vowels to a schwa in the plural base.

Base (singular)	Reduplicated (plural)	Gloss
kú:ɻaɲ	kúkəɻaɲ	medicine
má:wit	máməwit	lion
wí:nag	wípənag	sibling
má:kai	máməkāĭ	medicine man
má:gina	máməgina	car/machine (Sp)
ká:ɻit	kákə̱ɻit (wagon)	wagon (Sp)

TABLE 6. Reduction of full vowels to a schwa and shortened vowels in the reduplicant.

Previous literature does not mention the presence of the centralized vowel in the context of reduplication (cf. Fitzgerald 1999, Haugen 2009, Riggle 2006—on Pima, a mutually intelligible dialect, Zepeda 1983, Hill and Zepeda 1998). Instead, such reduplicated forms are presented with full syncope, that is, the loss of a word internal sound (discussion on syncope to follow). However, the contrast between words with reduction and syncope is unambiguously evident in Mr. Cruz's speech and is likely due to processes of language change, a topic which is given further treatment at the end of this section.

The final reduction process noted in our data also affects the plural form of the base, where the initial CV syllable in the singular base undergoes syncope (i.e. the vowel is absent), as

has been noted by others (e.g. Fitzgerald 1999, 2003; Riggle 2006—on Pima). This process additionally results in resyllabification leading to a closed initial syllable in the reduplicated form of the word (Fitzgerald 1999, 2003). In other words, initial consonants appearing in the base become the coda of the syllable comprising the reduplicant as represented in Table 7 (codas are bolded and '_' denotes syncope).

Base (singular)	Reduplicated (plural)	Gloss
gáso	gá**k**_so	fox
ʧʊ́so	ʧʊ́**ʧ**_so	chipmunk
kótoɲ	kó**k**_toɲ	shirts (Sp)
wísiɻo	wí**p**_siɻo	calf

TABLE 7. Vowel syncope in the reduplicated base.

In regards to the prosodic morphology hypothesis (McCarthy & Prince 1995), each derived (i.e. plural) word consists of a light reduplicant occurring with modified plural bases which do not necessarily pattern as heavy, however, the original singular forms of the bases do. The resyllabified plural bases are likely driven by principles of language change.

Syncope as well as the appearance of the schwa is presumably a result of the pervasive tendency in TO for strong, word-initial primary stress. In general, primary stress is also cross-linguistically known to lead to developments such as vowel reduction and possibly syncope in adjacent unstressed syllables (e.g. Bybee et al. 1998:274). Finally, the evolution towards syncope can also lead to the creation of an enhanced prominently stressed syllable (Fitzgerald 2003). Syncope would not occur, however, if the process would run counter to the phonotactics of the language. For example, laryngeal codas rarely (if ever) occur in TO as illustrated in Table 8 (unattested codas are bolded). As in the other cases, quantitatively complementary reduplication is maintained throughout the data set below (light reduplicants with heavy bases).

Base (singular)	Reduplicated (plural)	Unattested	Gloss
hódaĭ	hóhodaĭ	*hó**h**_dai	rock/stone
húdʒit	húhudʒit	*hú**h**_dʒit	lizard
hóʔi	hóhoʔi	*hó**h**_ʔi	thorn
ha:waɲ	háhawaɲ	*ha**h**_waɲ	raven/crow
ha:ʂaɲ	háhaʂaɲ	*há**h**_ʂaɲ	saguaro cactus

TABLE 8. Reduplication in the context of laryngeal onsets.

From a diachronic perspective, we hypothesize that as a result of prominently stressed word-initial syllables in TO, a potentially large set of words in the language are evolving from full vowels towards syncope with the centralization of vowels as an intermediary phase. This also leads to the resyllabification of plural bases noted in reference to examples of syncope shown in Table 7. We intend to further investigate this topic in a future study.

3. APPLICATIONS IN THE L2 CLASSROOM. In the previous sections, we demonstrated how McCarthy and Prince's (1995:334) quantitative complementarity principle is helpful in explaining a greater number of patterns found in plural reduplicative forms in TO than the syllable-copying account. The latter is the description used in the only teaching grammar

available in the language (Zepeda 1983). In TO, our data illustrate that in many cases of plural reduplication, light bases occur with heavy reduplicants and heavy bases occur with light reduplicants. Given that TO is an endangered language, having a unified description that explains a greater number of reduplication patterns is particularly useful in teaching TO to L2 learners. Although the principle of quantitative complementarity cannot account for every instance of reduplication in TO, presenting students with a way of understanding more types of reduplication patterns makes their knowledge of the language more complete. Additionally, it allows them to produce correct forms more frequently. We briefly describe how our analysis of TO reduplication processes can be applied in the classroom.

In order to teach students to apply the principles of quantitative complementarity in reduplicative plural forms, they would first need to learn the following fundamental concepts:

1. the difference between consonants and vowel sounds
2. light versus heavy syllables
3. light versus heavy reduplicants
4. long versus short vowels

Once students have learned to identify consonants and vowels and have become familiar with syllable structures and principles of weight in TO, they will have gained the knowledge they need to understand the general concepts behind quantitative complementarity (i.e. applying light reduplicants to heavy bases and vice versa) without necessarily using the formal linguistic terms associated with the theory in lesson plans. Students could be introduced to the concepts associated with quantitative complementarity with a worksheet that includes several different types of basic paradigms of singular and plural forms. Below is an abbreviated example of the possible contents of a worksheet.

PLURAL PATTERNS IN TOHONO O'ODHAM (SAMPLE WORKSHEET)[7]
Based on the data below, answer questions 1–4.

1. Compare the singular and plural forms for group (A). How is the plural formed? Can you fill in what you think the plural form would be for *makai* 'doctor'?

(A) SINGULAR	PLURAL	ENGLISH TRANSLATION
nakʃɨɭ	nanakʃɨɭ	*scorpion*
ʧeodʒ	ʧeʧeodʒ	*boy/man*
ʔuwi	ʔuʔuwi	*woman*
makai	_____	*doctor*

2. Are there any differences in the way the plural is formed in (B) below compared to (A) above? What do you think the plural form would be for *ʈad* 'foot'?

(B) SINGULAR	PLURAL	ENGLISH TRANSLATION
ban	ba:ban	*coyote*
mad	ma:mad	*child (of a woman)*
dʒɨg	dʒɨ:dʒɨg	*hole*
ʈad	_____	*foot*

3. Again, look at the singular and plural forms. Do you note any differences in the patterns in (C) compared to (A) and (B) above? Try to fill in the answer for *mi:stoɭ* 'cat'.

(C) SINGULAR	PLURAL	ENGLISH TRANSLATION
ma:kai	mamakai	*medicine man*
ɭi:wa	ɭiɭiwa	*jacket*
ma:wit	mamawit[8]	*lions*
mi:stoɭ	_____	*cat*

4. Revisit groups (A)–(C) above and the observations you noted about the parts of the words that are repeated to create the plural. Can you come up with possible explanations for the different patterns?

The leading questions in the worksheet encourage the students to try and familiarize themselves with the plural formations and to identify the plural patterns in TO as if they were a puzzle to be solved without explicitly being told what the patterns are.

As a follow-up listening exercise to the worksheet, a teacher could create a version of Bingo as one possible activity to reinforce reduplication patterns in TO.[9] In short, students would receive cards with pictures of objects (as opposed to numbers in the traditional format of the game), possibly depicting some of the words in the worksheet proposed above. A portion of the pictures would show singular objects, for example, one ɻiꞏwa 'jacket' and others would display plural objects, such as two (or more) ɻiɻiwa 'jackets'. Such an activity would familiarize the students with new vocabulary items and with a basic understanding that the plural in TO is formed by reduplication. If teachers wanted to go further and have their students identify when short versus long reduplicants are used, they might want to ask students to write down what they think the reduplicated form looks like based on the structure of the singular form as each word is called out. If the class is at a very introductory level, the teacher might want to provide a list of the relevant words in the singular forms, so the students could refer to the list for assistance in determining if a short or long vowel should be used in the reduplicant. If a student believes that s/he has won the game, the teacher could ask the student to recite each of the relevant items on the person's card, practicing the production of various singular and plural forms.

Finally, there are dialectal variants of TO that do not necessarily adhere to all of the forms shown in the worksheets. As this issue arises in this classroom, this would be an opportunity to discuss with students concepts related to language and dialectal variation, emphasizing that no form is more correct than another. For example, we have heard the following variations in reduplication among various TO speakers:

místoɻ   →   mí**mi**stoɻ   OR   mí**m**_stoɻ

4. CONCLUSIONS AND FUTURE RESEARCH. The only pedagogical grammar on TO (Zepeda 1983) conflates multiple reduplication processes into a single pattern—CV-copying—likely in an effort to maintain pedagogical simplicity for a complex process. However, our application of McCarthy and Prince's (1995:334) theory of quantitative complementarity augments the reduplication patterns that can be explained in a straightforward manner in an L2 setting. In sum, the shape of the reduplicant is largely contingent on the weight of the base. CVV reduplicants appear when the base is light (CVC) and CV reduplicants occur when the base is heavy (e.g. CVCV, CVVCV, etc.). In some cases, this results in lengthening or shortening of the reduplicant in order to maintain a complementary weighted relationship with the base.

The concept of quantitative complementarity, although theoretical in nature, is useful in teaching practice. A practical application of the concepts makes it possible to teach L2 learners of TO a larger number of reduplication patterns while maintaining pedagogical simplicity and efficacy. It also enhances recall and adds explanatory power to the description of the data by providing an explanation for why the reduplicants have the forms that they do.

In future research, we wish to examine other varieties of TO in order to see whether quantitative complementarity is generally applicable across varieties. Also, we hypothesize that the reduction process observed in plural bases (cf. Tables 6 and 7) is a prominence enhancement phenomenon, which is driven by TO's strong, word-initial primary stress. Based upon our preliminary observations, we expect that different dialects will be at various stages in the reduction process affecting plural base forms. Cross-dialectal data showing plural bases at

different stages of reduction could in turn inform us on how these processes of language change affect syllable structure and possibly even the shape of the reduplicant (i.e. heavy or light). Finally, future research might shed light on the ramifications of language endangerment on reduplication processes (as well as on other patterns in the language) especially among younger speakers who may not have the opportunity to fully acquire the language or do not use TO on a regular basis.

REFERENCES

BYBEE, JOAN; PAROMITA CHAKRABORTI; DAGMAR JUNG; and JOANNE SCHEIBMAN. 1998. Prosody and segmental effects: Some paths of evolution for word stress. *Studies in Language* 22.267–314.

FITZGERALD, COLLEEN. 1999. Unfaithful bases and syncope in Tohono O'odham reduplication. Paper presented at the West Coast Conference on Formal Linguistics 18, Phoenix.

FITZGERALD, COLLEEN. 2003. How prosodically consistent is Tohono O'odham? *Studies in Uto-Aztecan* 5.55–74.

FITZGERALD, COLLEEN. 2010. Language documentation in the Tohono O'odham community. *Language Documentation: Practice and values*, ed. by Lenore A. Grenoble and N. Louanna Furbee, 231–240. Amsterdam: John Benjamins Publishing.

HAUGEN, JASON. 2009. Enhancement and diachrony in reduplication. *Grazer Linguistische Studien* 71.39–58.

HAYES, BRUCE. 1989. Compensatory lengthening in moraic phonology. *Linguistic inquiry* 20.253–306.

HILL, JANE, and OFELIA ZEPEDA, 1992. Derived words in Tohono O'odham. *International Journal Of American Linguistics* 58.355–404.

HILL, JANE, and OFELIA ZEPEDA. 1998. Tohono O'odham (Papago) plurals. *Anthropological Linguistics* 40.1–42.

MADSEN, KENNETH. 2004. *Tohono O'odham Nation comprehensive education study 2004*. Sells, AZ: Tohono O'odham Nation and Tohono O'odham Community College.

MCCARTHY, JOHN, and ALAN PRINCE. 1995. Prosodic morphology. *The handbook of phonological theory*, ed. by J. Goldsmith, 318–366. Cambridge: Blackwell Publishers.

RIGGLE, JASON. 2006. Infixing reduplication in Pima and its theoretical consequences. *National Language Linguistics Theory* 24.857–891.

SAXTON, DEAN; LUCILLE SAXTON; and SUSIE ENOS. 1983. *Dictionary, Tohono O'odham / Pima to English, English to Tohono O'odham / Pima*. Tucson: The University of Arizona Press.

SAXTON, DEAN, and LUCILLE SAXTON. 1969. *Dictionary: Papago & Pima to English (O'odham—Mil-gahn); English to Papago & Pima (Mil-gahn—O'odham)*. Tucson: University of Arizona Press.

STUBBS, BRIAN D. 1995. The labial labyrinth in Uto-Aztecan. *International Journal of American Linguistics* 61.396–422.

TRANEL, BERNARD. 1991. CVC Light syllables, geminates and moraic theory. *Phonology* 8.291–302.

ZEPEDA, OFELIA. 1983. *A Papago grammar*. Tucson: University of Arizona Press.

NOTES

[1] All TO examples are given in IPA.

[2] DIST=distributive, ITER=interative, PUNCT=punctual, SG=singular, PL=plural, SUB=subject,

[3] For alternative analyses of long vowel reduplication, see Hill and Zepeda (1992, 1998), Fitzgerald (2003)

[4] The word-initial /w/ has been historically reconstructed as /p/ within the Uto-Aztecan family. In TO, the voiceless bilabial stop does not occur word-initially but is permitted word-medially (cf. Stubbs 1995).

[5] Word-final short vowels are voiceless unless followed by a voiced segment.

[6] Sp=Spanish

[7] The example worksheet is given in IPA instead of TO orthography for consistency with the rest of the paper.

[8] In §2.3 the plural for this word was shown with the schwa, but the schwa is present due to phonological processes and not because it is a part of the phonemic inventory. Such detail is unnecessary for pedagogical purposes.

[9] Various rules for how to play Bingo can be found at: http://www.wikihow.com/Play-Bingo

Prefabs and priming in second-person address
in New Mexico and Southern Colorado

Aubrey Healey
University of New Mexico

ABSTRACT. This study examines second-person singular address (2S) forms in New Mexican Spanish to discover the reasons that a speaker produces one form over another. Four hundred and eighty nine tokens of 2S verbs were extracted in their context from the NMCOSS (Bills & Vigil 2008). Multi-variate analysis was performed on linguistic and extralinguistic factors. The factors affecting 2S form choice are: formulaic expressions occurring with high frequency, linguistic priming, and the genre of discourse. This study is the first to examine 2S address form usage in New Mexican Spanish quantitatively by examining both linguistic and extra-linguistic factors.

Keywords: Second-person address, formulaic expressions, prefabs, Spanish, New Mexico, variation, priming, genre effects

1. INTRODUCTION. It is common knowledge that there are multiple forms of address for second-person singular (2S) in Spanish, specifically, *tú, vos,* and *usted.* Recently, I investigated second-person address form usage in Cali, Colombia with the aim of comparing form usage across regions (Healey 2011). The current research comes at an opportune time as the Spanish of New Mexico and Southern Colorado has become a popular topic of late in the field of sociolinguistics due to the fascinating nature of the contact between Spanish and English in this area (c.f. Bills & Vigil 1999, 2008, Torres Cacoullos & Travis 2011). The *vos* form is very uncommon in this dialect, thus this study will only be comparing usage of *tú* and *usted.* Examples 1 and 2 below demonstrate the nature of the variation found in second-person address form usage in New Mexico and Southern Colorado (NMCOSS, Bills & Vigil 2008).

(1)   *..Si **recuerdas** algo de .. curanderos o curanderas, si **recuerdas** algunos … cuentos de brujería, .. si -- **recuerdas** algunas cosas de la muerte, de velorios, o de funerales.*

   '.. If **you (*tú*) remember** something about healers (m.) or healers (f.), if **you (*tú*) remember** some … stories about witchcraft, .. if -- if **you (*tú*) remember** some things about death, about wakes, or funerals.'

<div align="right">(4-2A2, lines 9-15)</div>

(2)   *Cuando= .. **usted estaba** niña, qué le platicaban de la historia de= -- .. de= Colorado y de Nuevo México?*

   'When= .. **you (*usted*) were** a girl, what did they tell you about the history of= -- .. of= Colorado and New Mexico?'

<div align="right">(10-8A, 109-111)</div>

This study examines second-person address (2S) forms in New Mexican Spanish with the aim of discovering linguistic and extralinguistic reasons that a speaker chooses one form over another. Previous research on Spanish has found social factors, involving the dynamics of power and

solidarity, to be the strongest influence in 2S form choice (Brown & Gilman 1960, Plasencia 1997, Solé 1978. Uber 2000). Interestingly, my recent study (Healey 2011) on 2S forms in Cali, Colombia found linguistic and cognitive factors, such as occurrence in prefabricated constructions (c.f. Bybee 2003, 2006, Travis 2006, Wilson 2009) and linguistic priming (c.f. Torres Cacoullos & Travis 2011, Travis 2007), to be more significant than social factors.

Comparing the New Mexican data with my previous *tú, vos* and *usted* study of data from Cali, Colombia will not yield equal results because the genre of discourse is different for each study (Cali: spontaneous conversation vs. New Mexico: sociolinguistic interview), and the status of Spanish in each place is different. In Cali, Colombia the population is generally monolingual in Spanish, whereas the New Mexican data has patches of code-switching, which is considered acceptable and often goes unnoticed in relaxed speech. However, the usage of prefabricated constructions and structural priming are the result of cognitive processes that are common to all humans, no matter the language that they speak (Bybee 2003, 2006, Croft & Cruse 2004, Goldberg 2003, Fillmore et al. 2006, Travis 2007, *inter alia*). Thus, I expect to find those processes at work in these data as well. In the following sections, this paper will show that structural priming, the use of formulaic expressions, and the genre of discourse significantly affect a speaker's form choice when addressing an interlocutor.

2. THEORETICAL FRAMEWORK

2.1. EXTRALINGUISTIC FACTORS. Previous literature regarding 2S pronoun usage in Spanish pointed to social factors being the determinants of 2S form (Brown & Gilman 1960, Plasencia 1997, Solé 1978, Uber 2000). Brown and Gilman (1960) noted that the most relevant factors affecting 2S form are social, specifically that the choice between forms is based upon the speaker's perception of the dynamics of power and solidarity between themselves and the addressee. Speaking on the variation of 2S forms in Puerto Rico, Uber stated that "Power may be manifested by differences in age, profession, or perceived position between the speakers involved … Solidarity can refer to the degree of familiarity or intimacy between speakers" (2000:316), and in this way operationalized this notion for her study. The fascinating nature of the 2S form in Spanish also stems from the fact that most Spanish-speakers are aware of the social connotations expressed by using each form as conventionalized in the speech community. They have internalized the social norms associated with these forms as metalinguistic knowledge, which is why it is often difficult to explain to a native Spanish-speaker that the formal/informal distinction does not exclusively dictate their use. In the NMCOSS data many interviewees are retired and all of the interviewers are students, so the only social factors that this study tested for were gender, age and speaker role as interviewee or interviewer. It should be noted that although all of the interviewers are students they are not all in the same age group, and the same can be said for the interviewees—some are elderly and some are middle aged or young adults. This is relevant because it shows that the factor of speaker role is not confounded with the factor of speaker age. The factors I coded for are discussed in section 4.

2.2. LINGUISTIC AND COGNITIVE FACTORS. In general, the previous research on pronouns and subject forms is divided into two schools: those who believe that the driving force behind form use is social (this mostly applies to 2S research) and those who believe it to be linguistic. In recent years it has become accepted that language is constantly emerging through use (Bybee, 2003, 2006, DuBois 2003, Hopper 1998, *inter alia*); which is to say that language is used for interaction. It is the different uses and the context in which the forms occur that change their meaning and form, and even the grammar of a language itself over time. Linguists who follow this theory have noted the way in which prefabricated constructions, or prefabs (Bybee 2003, 2006, Wilson 2009), tend to show little variation in form. According to the theory of Emergent Grammar (DuBois 2003, Hopper 1998, Bybee et al. 1994, *inter alia*), all language is behavior learned through repetition (Bybee 2003, 2006) and is made up of constructions which can be more or less lexically filled (Croft & Cruse 2004, Goldberg 2006, Fillmore et al. 2003, *inter alia*). The current study will investigate the way prefabricated constructions affect second person form choice based on this framework.

According to Torres Cacoullos and Travis (2011:3), "the linguistic factors consistently found to have an effect of the patterning of subject expression are switch reference, semantic class of verb, Tense-Mood-Aspect and person, while social factors such as age, gender and socioeconomic status are typically found to have no effect." In my study of 2S forms in Cali (Healey 2011), I also found linguistic factors to more significantly affect form choice, and since there were no other studies of 2S form which use multivariate analysis, it was decided that the same group of factors should be coded for in this study. Travis (2007) found that the genre of the data can often affect the rates of subject expression in Spanish. She found higher rates of *yo* expression in conversation than in personal narrative, but identical patterning. Specifically, the narratives had a high degree of subject continuity, while interactive conversation had much more topic-shifting and less subject continuity. This is important for the current study of *tú* and *usted* because second person address forms are used for interaction and will thus be much more common in spontaneous conversation than in narrative. In fact, a second hypothesis of this study is that the interview setting has affected the speakers' second person address form choice. This study advances previous research on second person address forms by considering the set of factor groups that may account for the variation together in multivariate analysis (Sankoff et al. 2005).

3. HYPOTHESES

(1) Certain formulaic expressions (prefabs) occur with such high frequency that they tend to be uttered in the same second person form.
(2) Second person form choice will be affected by priming, which means that the previous use of *tú* will prime the current use of *tú*, and the previous use of *usted* will prime the current use of *usted*.
(3) The genre of discourse will affect the overall patterning of second person address forms. In particular, because second person address forms are much more common and useful in interactive rather than narrative discourse, the linguistic and extralinguistic sub-contexts affecting 2S form choice will be easier to discover in spontaneous conversation, whereas the speaker and addressee roles found in the interview genre may obscure the effects of the more subtle influences.

4. DATA AND METHODOLOGY. In what follows, the above hypotheses were tested by applying the variationist method to the New Mexican and Southern Colorado Survey of Spanish (NMCOSS) (Bills & Vigil 2008). The key assumption of the variationist method is that discourse is characterized by a form-function asymmetry (Labov 1969, Torres Cacoullos & Travis 2011). For this study, this means that there is variation between the 2S address forms *tú* and *usted* in this community, even though they both serve relatively the same function of addressing the speaker's interlocutor. However, although they both serve the same function, the usage of these forms is conditioned by the exact context that they occur in. This variability is structured and conditioned by both linguistic (cognitive and structural factors) and extralinguistic factors (sociodemographic factors) as they affect usage in context (Torres Cacoullos & Travis 2011, Travis 2007). Therefore, we must observe the linguistic and extralinguistic conditioning of forms in variation in order to reveal the structure of discourse. This will show us which factors tend to co-occur, the relative frequency of their co-occurrence, as well as the linguistic context in which they occur (Torres Cacoullos & Travis 2011). Thus, I am able to test the above hypotheses by operationalizing elements of the context into factors which can be examined in multivariate analysis to discover their interaction and relative magnitude of effect.

4.1. CODING. To identify the linguistic conditioning of 2S usage, 489 tokens of second person singular verbs were extracted from the NMCOSS (Bills & Vigil 2008) in their discourse context, excluding truncated tokens. All tokens were coded for the following linguistic factor groups: current 2S form (the dependent variable), 2S form (priming, both within the same speaker and across speakers, with no limit on Intonation Units (IUs) provided that there were no other 2S forms in between), realization (whether the pronoun was used with the verb or not), which verb it is (to test for formulaic expressions, if a particular verb (e.g. *mirar* 'to look')had more than 50% of its uses in the same 2S form), semantic class of verb, clause type, reported speech, and Tense-Mood-Aspect. These groups represent linguistic sub-contexts that are likely to affect the variability of form usage, and are a means to operationalize hypotheses about the constraints affecting second person form usage.

Additionally, I coded for a number of extralinguistic factors: speaker role (interviewer/interviewee), speaker age, addressee age, speaker gender, and addressee gender. These extralinguistic factors were chosen due to past research on 2S forms that also used the sociolinguistic interview as the genre of discourse for analysis (Brown & Gilman 1960, Uber 2000, *inter alia*). All factor groups were simultaneously analyzed in Variable-rule analysis using GoldVarb X (Sankoff et al. 2005). The multiple regression procedure in Variable-rule analysis works to determine the factor groups which interact together and as a group account for the largest amount of the variation being studied in a statistically significant way (Sankoff 1988, Torres Cacoullos & Travis 2011).

5. RESULTS

5.1. VARIABLE-RULE ANALYSIS. Table 1 presents the results from a Variable-rule analysis of factors contributing to the speaker's 2S form choice. The input (.779) indicates the overall likelihood that the variant (*usted*) will occur (the overall rate of *usted* use was 66.6%). The factor groups that had a significant effect on 2S form choice were: speaker role, previous 2S form in the discourse, and realization. Factors not listed in the table were not significant. Within each factor group, the ordering of the factor weights shows the direction of effect, which tells us which

factors favor/disfavor *usted* usage, based on whether the factor weight is over (favor) or under (disfavor) 0.5. We can see that the factors which favor *usted* usage were: if the speaker is an interviewer, if the previous 2S form used was *usted* or if there was no previous form at all (like at the beginning of a conversation, which was usually started by the interviewer), and if the current form is expressed. The factors which disfavor *usted* usage (and in this study would thus favor *tú* usage) were: if the speaker is an interviewee, if the previous 2S form used was *tú*, and if the current form is unexpressed. The range is the difference between the highest and lowest factor weight in each factor group, which indicates the relative strength (referred to as the magnitude of effect) of each factor group (Torres Cacoullos & Travis 2011). We can see in the table that the speaker role has the largest magnitude of effect on *usted* usage, followed by previous 2S form and realization, respectively. The last column in the table shows the percentage of *usted* usage in each environment.

N=489	Input= .779 (66.6% usted, 326/489)	Log likelihood= -157.190

Factor	Factor Weight	Percent *usted* (n)
Speaker Role		
Interviewer	.790	93% (284/304)
Interviewee	.102	23% (42/185)
Range	*.69*	
Previous 2S Form		
Usted	.629	82% 258/315)
None	.513	84% (16/19)
Tú	.253	34% (52/155)
Range	*.38*	
Realization of Pronoun		
Expressed	.635	80% (111/139)
Unexpressed	.445	61% (215/350)
Range	*.19*	

TABLE 1. Variable-rule analysis of the contribution of factors selected as significant to *usted* usage p<.05 (p =0.032).

Each significant factor group will now be discussed in detail. For the first factor group, speaker role, *usted* was more likely to be used by the interviewer (factor weight .79) than by the interviewee (.10). This makes sense because the extralinguistic sub-context created by the interview genre creates a "dynamic of power and solidarity" (Brown & Gilman 1960, Uber 2000) which triggers the *usted* usage by the interviewer to show respect. Previous studies (Brown & Gilman 1960, Plasencia 1997, Solé 1978, Uber 2000) have found social factors such as speaker and addressee age, gender, and relative social status to trigger one 2S form over the other. These factors were not found to be significant in this study, which demonstrates the strength of the power dynamic in the interviewer-interviewee interaction. The status of the addressee as the interviewee seems to trump the effects of the other extralinguistic factors.

The second factor group selected as significant was previous 2S form. *Usted* was more likely to be chosen when the previous 2S form used was also *usted* (factor weight .63) than when the previous 2S form used was *tú* (.25). This can be seen in the following examples, where we see a previous *usted* usage favoring continued *usted* use within the same speaker (as in example

3) and a previous *usted* usage favoring continued *usted* usage across two different speakers (as in example 4).

(3)   A: *.. Ahora vamos a --*
           Este=,
           dígame *.. dónde* **nació usted**.

       A: .. Now we're gonna --
           Um=,
           Tell me (*usted*) .. where **were you (*usted*) born**.

                                                                    (20-1A1, 61-63)

(4)   *A: (H)Y este=,*
           usted *ah* **trabajaba** *cuando= ..* **estaba** *más joven,*
           [XXX].
       R: [Eh=,
           yo trabajé],
           por mi esposo.
           Siempre.
           siempre.
           Él --
           como le digo,
           él --
           él manejaba companías de tresquila. ((COMPAÑÍAS; TRASQUILA))
       A: Ah=.
       R: *No sé si* habrá oído decir usted *de --
           de tresquila.

       A: (H) And um=,
           You (*usted*) ah **were working** when= .. **you (*usted*) were** younger,
           [XXX].
       R: [Eh=,
           I worked],
           for my husband.
           Always.
           always.
           He --
           like I'm telling you,
           he --
           he managed shearing companies.
       A: Ah=
       R: I don't know if **you'll (*usted*) have heard** of --
           of shearing.   (20-1A1, 625-639)

The effect of the previous 2S form on the current 2S form choice is known as structural priming. Structural priming is the cognitive effect "whereby the use of a certain structure in one utterance functions as a prime on a subsequent utterance, such that that same structure is repeated" (Torres

Cacoullos & Travis 2011:13). Priming in Colombian Spanish was also discovered in my previous study of 2S address forms (Healey 2011), as well as for 1S subject expression (Travis 2007). An example of structural priming across two speakers in Colombian Spanish is given below in 5.

(5)   *Ángela: [2Pero **tú estabas**2] des- --*
            ***tú no te acuerdas**?*
        Santi: .. Pues,
            *Yo me acuerdo que sí me **pasaste**,*
            pero no me acuerdo a qué horas,
            ni nada.

        Ángela: [2 But **you** (*tú*) **were**2] awak- --
            You (*tú*) **don't remember**?
        Santi: .. Well,
            I remember that **you** (*tú*) **passed** me,
            but I don't remember what time,
            or anything.

(Travis 2005, Almuerzo, 435-440)

It can be seen from examples 3-5 that priming has a definite effect on 2S form choice. In Healey (2011), priming had the largest magnitude of effect, but in the current study the priming effect was overshadowed by the effect of speaker role. This is evidence that the genre of the discourse can affect the results of a study such as this.

        The results for the third factor group, realization, indicate that *usted* usage was favored when the speaker chooses to express the 2S form with a pronoun (factor weight .64) rather than leave it unexpressed (.44). However, this factor group had the smallest magnitude of effect (range .19). This could also be a side-effect of the genre of the discourse because the interviewers were more likely to use *usted* out of respect for the interviewee and because the 3S form is ambiguous between *él, ella* and *usted* when used without a pronoun. In a context where *usted* would tend to be used, priming of *tú* may override the social norms dictating *usted* use. However, in cases where the social factors and linguistic factors are in competition priming will take effect unless the speaker has (perhaps unconsciously) decided to use a particular form with their interlocutor all the time, like the situation where an interviewer decides to use *usted* to address their interviewee. It is evident that the interviewers chose to use *usted* with the interviewees because they ask some specific questions in the *usted* form (e.g. *¿Cuándo aprendió usted español?* When did you learn Spanish?') that are the same for all interviewees. As we will see in section 5.2., the use of certain prefabricated constructions can trump both the social and the linguistic factors affecting 2S form choice.

5.2. PREFABRICATED CONSTRUCTIONS. It is vital to remember that some tendencies that emerge as significant in the data as a whole may be skewed by their occurrence in prefabricated constructions (prefabs) (Bybee 2003, 2006, Company Company 2006, Erman & Warren 2000, *inter alia*). Healey (2011) found several prefabs where the specific 2S form used in the expression made up more than 50% of the instances of that verb. The prefabs discovered in that study were: *mira* ('look'), *venga* ('come here'), *oiga* ('listen'), *ve* ('look'), *fíjate* ('watch out/check it out/look'), and *imagínase* ('just imagine'). The first thing to notice about these prefabs is that they are all imperatives and also almost all discourse markers used to take the floor. Thus, I expected to find imperative prefabs for these same verbs in New Mexican Spanish since they serve such a general discourse function.

For the current study, the following formulaic expressions were found (meaning that 50% or more of the instances of these verbs occurred in this form): *oiga* ('listen'), *mira* ('look'), *habla* ('you speak/talk'), *aprendió* ('you learned'). These are used despite previous uses of a different 2S form. However, *habla* and *aprendió* probably only appear to be prefabs because of the genre and subject matter; interviewers were required to ask the interviewees how they learned Spanish and some things about their speech. In fact, only 12.5% (3/24) of the *aprender* tokens are produced by interviewees and none of them are in the form *aprendió*. Likewise, only 15% (5/34) of the *hablar* tokens are produced by interviewees and none of them are in the form *habla*. Thus, the only *real* prefabricated constructions with specific discourse functions found in this study were: *oiga* with 68% (21/31) of the tokens of *oír* occurring in this form and *mira* with 60% (12/20) of the tokens of *mirar* occurring in this form. Examples of these prefabs are given in 6 and 7 below.

(6)   *D: ... **Oiga**,*
          y este=,
          y ahora qué hace?
          En qué trabaja?

        D: … **Listen (usted)**,
          and this guy=,
          and now what does he do?
          what does he do for work?

                                                                    (102-1A1, 327-330)

(7)   ***P:** [No quieren].*
          No quieren.
          *No quieren **oiga**.*
          % Dicen,
          [Pueden llevar] --
        A: [I didn't want it] either.
        P: pueden --
          pueden llevar al caballo a beber agua,
          but that don't mean he's gonna drink.
        A: But if that water is not .. appropriate for our kids,
        P: <X Pos luego no X> --
          Hay --
          como dijimos,

hay chavalos que sí quieren.
y otros que di- --
no más se= --
A: *Lo que pasa es que cuando van a la escuela,*
   tienen que dejar su cultura y su lengua.
P: *[Pos] --*
A: *[like] you and I?*
P: Pos **mira**,
   yo tengo --
   .. my nephew ahora,
   .. y es suerte.

P: [They don't want to].
   They don't want to.
   They don't want to **listen (*usted*)**.
   % They say,
   [They can bring] --
A: [I didn't want it] either.
P: they can --
   they can bring the horse to water,
   but that don't mean he's gonna drink.
A: But if that water is not .. appropriate for our kids,
P: <X well later no X> --
   There's --
   like we said,
   there's kids that do want to.
   and others that sa- --
   not more than= --
A: What happens is when they go to school,
   they have to leave behind their culture and their language.
P: [Well] --
A: [like] you and I?
P: Well **look (*tú*)**,
   I have --
   .. my nephew now,
   .. and he's lucky.

(88-1A3, 654-677)

As we can see from these examples, *oiga* and *mira* are clearly prefabs not only because these instances make up more than 50% of the tokens of their verbs, but also because their lexical meaning is mostly bleached in these instances and instead they have a specific function, that of being discourse markers used to take the floor and draw attention to what the speaker is about to say (c.f. Travis 2005 for a study of discourse markers in Colombian Spanish). Discourse markers have been shown to be "regulatory units," which control the flow of the conversation (Chafe 1994). The speakers in the above examples 6 and 7 are not actually telling their interlocutor to listen or look, but instead they are merely grabbing their attention and directing it to their next

utterance. Even stronger evidence that these are prefabs is the fact that in example 7 the same speaker uses both *oiga (usted)* 'listen' and *mira (tú)* 'look' to address the same interlocutor, showing that these discourse markers are mostly fixed in their 2S form and function, one in the *tú* form and one in the *usted* form. Because I also found the phrases *oiga* and *mira* in my previous study of monolingual Colombian Spanish, it is reasonable to propose that *oiga* and *mira* are prefabs used as discourse markers in many varieties of Spanish.

6. SUMMARY AND CONCLUSION. In sum, the linguistic impact of structural priming and formulaic expressions has significant effects on second person form chosen by a speaker. The effect of structural priming is shown by the finding that a previous *usted* usage was much more likely to trigger current *usted* usage than if the previous 2S form was *tú,* both within one speaker and across speakers. The impact of formulaic expressions on second person form choice is that certain expressions, in particular *oiga* 'listen' in the *usted* form and *mira* 'look' in the *tú* form, will tend to occur specifically in these 2S forms and with particular discourse functions. In particular, these prefabs are discourse markers used to grab the interlocutor's attention and direct it to what the speaker is about to say. Also, the discourse markers *oiga* and *mira* may not follow priming effects or social 2S norms, and thus *oiga* may be used with someone to whom the speaker was just using the *tú* form, or with whom they normally use the *tú* form, and likewise *mira* may be used to grab the attention of an interlocutor with whom the speaker normally uses the *usted* form even if they just used the *usted* form with them.

Thus, the forms *oye (tú)* and *mire (usted)*, although they mean the same thing lexically as *oiga (usted)* and *mira (tú)*, are less likely to occur in this dialect for these discourse functions due to the automated cognitive procedures controlling these processes (c.f. Bybee 2003, 2006). The effects of priming and prefabs are cognitive in nature, with both aimed at reducing the mental work involved in interaction, and therefore these effects will be found both across different varieties of Spanish, and also cross-linguistically. The fact that my previous work (Healey 2011) on second person address forms in Colombian Spanish also found priming and prefabs to be statistically significant factors affecting 2S form usage corroborates this finding.

However, the key difference between my previous study and the current study is the genre of discourse. My previous study was based on spontaneous conversational data, whereas the current study is based on sociolinguistic interviews with long stretches of personal narrative. Interestingly, this genre difference was quite important. As was shown in the above sections, the most significant factor affecting 2S form choice in these data was whether the speaker was an interviewer or interviewee. An interviewer was much more likely to address the interviewee with the *usted* form than the *tú* form. This finding can be attributed to the "dynamics of power and solidarity" (Brown & Gilman 1960, Uber 2000) because in the (semi)formal setting of an interview the interlocutors are more likely to follow the social norms of their roles, with the interviewer showing respect to the interviewee by using the *usted* form with her/him. This is demonstrated by the large magnitude of effect exhibited by the factor group of speaker role. In future studies of second person address forms in New Mexico, it would be better to use spontaneous conversational data, where the regular patterns of natural discourse are most easily observed (DuBois et al. 1993, DuBois 2003, Torres Cacoullos & Travis 2011, *inter alia*). There are more opportunities for speaker change and second person address use between interlocutors in a conversation than in an interview eliciting large chunks of narrative; thus, it is logical that speaker roles would turn out to be the most significant factor of this study.

REFERENCES

AMMON, ULRICH; NORBERT DITTMAR; and J. KLAUS MATTHEIER (EDS.). 1988. *Sociolinguistics: An international handbook of the science of language and society* (vol. 2). Berlin/New York: Walter de Gruyter.

BILLS, GARLAND, and NEDDY VIGIL, 1999. Ashes to ashes: The historical basis for dialect variation in New Mexican Spanish. *Romance Philology* 53.43–67.

BILLS, GARLAND, and NEDDY VIGIL. 2008. *The Spanish language of New Mexico and southern Colorado: A linguistic atlas.* Albuquerque: University of New Mexico Press.

BROWN, ROGER, and ROGER GILMAN. 1960. The pronouns of power and solidarity. *Style in language*, ed. by Thomas A. Sebeok , 253–276. Cambridge, MA: Massachusetts Institute of Technology Press.

BYBEE, JOAN. 2003. Mechanisms of change in grammaticalization: the role of frequency. *Handbook of historical linguistics*, ed. by Brian D. Joseph and Richard D. Janda, 602–623. Oxford: Blackwell Publishers.

BYBEE, JOAN. 2006. From usage to grammar: The mind's response to repetition. *Language* 82.711–733.

BYBEE, JOAN; REVERE PERKINS; and WILLIAM PAGLIUCA. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.

CHAFE, WALLACE. 1994. *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

COMPANY COMPANY, CONCEPCIÓN. 2006. Subjectification of verbs into discourse markers: Semantic pragmatic change only? *Belgian Journal Of Linguistic* 20.97–121.

CROFT, WILLIAM, and D. ALAN CRUSE. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.

DUBOIS, JOHN; STEPHEN SCHUETZE-COBURN; SUSANA CUMMING; and DANAE PAOLINO. 1993. Outline of discourse transcription. *Talking data: Transcription and coding in discourse research*, ed. by Jane A. Edwards and Martin D. Lampert, 45–89. Hillsdale, NJ: Lawrence Erlbaum Associates.

DU BOIS, JOHN. 2003. Discourse and grammar. *The new psychology of language: Cognitive and functional approaches to language structure* (vol. 2), ed. by Michael Tomasello, 47–88. New Jersey: Lawrence Erlbaum.

ERMAN, BRITT, and BEATRICE WARREN. 2000. The idiom principle and the open choice principle. *Text.* 20.29–62.

GOLDBERG, ADELE E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

FILLMORE, CHARLES; PAUL KAY; and MARY CATHERINE O'CONNOR. 2003. Regularity and idiomaticity in grammatical constructions: The case of *let alone. The new psychology of language: Cognitive and functional approaches to language structure*, ed. by Michael Tomasello, 243–270. Mahwah, NJ: Lawrence Erlbaum Associates.

HEALEY, AUBREY. 2011. Tú, vos, or usted: Who are you in Cali?. Paper presented at The Linguistics Society of the Southwest Annual Conference, University of Texas at Brownsville.

HOPPER, PAUL J. 1998. Emergent grammar. *The new psychology of language: Cognitive and functional approaches to language structure* (vol. 2), ed. by Michael Tomasello, 155–175. Mahwah, NJ: Lawrence Erlbaum Associates.

LABOV, WILLIAM. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45.715–762.

PLACENCIA, MARÍA ELENA. 1997. Address forms in Ecuadorian Spanish. *Hispanic Linguistics* 91.165–202.

SANKOFF, DAVID. 1988. Variable rules. *Sociolinguistics: An international handbook of the science of language and society* (vol. 2), ed. by Ulrich Ammon, Norbert Dittmar, and Klaus J. Mattheier, 984–997. Berlin;New York: Walter de Gruyter.

SANKOFF, DAVID; SALI TAGLIAMONTE; and ERIC SMITH. 2005. *Goldvarb X: A variable rule application for Macintosh and Windows*. Department of Linguistics, University of Toronto.

SOLÉ, YOLANDA RUSSINOVICH. 1978. Sociocultural determinants of symmetrical and asymmetrical address forms in Spanish. *Hispani*a 14.940–949.

TRAVIS, CATHERINE E. 2005. *Discourse markers in Colombian Spanish*. Berlin;New York: Mouton de Gruyter.

TRAVIS, CATHERINE E. 2006. Subjetivización de construcciones. *VIII Encuentro Internacional de Lingüística en el Noroeste, Siere Memorias* (vol. 2), ed. by Rosa María Ortiz Ciscomani, 85–109. Sonora, México: Unison.

TRAVIS, CATHERINE E. 2007. Genre effects on subject expression in Spanish. *Language Variation and Change* 19.101–135.

TORRES CACOULLOS , RENA, and CATHERINE E. TRAVIS. 2011. Testing convergence via code-switching: Priming and the structure of variable subject expression. *International Journal of Bilingualism* 15.241–267.

UBER, DIANE RINGER. 2000. 'Addressing' business in Puerto Rico: *Tú* vs. *usted*. *Research on Spanish in the United States: Linguistic issues and beyond*, ed. by Ana Roca, 310–318. Somerville: Cascadilla Press.

WILSON, DAMIÁN VERGARA. 2009. Prefabs at the center for centuries; The persistence of formulaic sequences in the Spanish change-of-state construction quedar(se) + adjective. Paper presented at the 12th Hispanic Linguistics Symposium (HLS). San Juan, PR.

Two speech communities in Puerto Rico: An ethnographic study about social class and children learning English in public and private schools of the island.

Jannette Hermina
University of New Mexico

ABSTRACT. This paper describes an ethnographic study related to the learning of English in a public school and a private school in two small towns on the northern coast of Puerto Rico. The research examines the social interaction of elementary school students in the English classroom, and the cultural capital that families use to increase the learning of English in their children. Cultural capital is a term coined by Pierre Bourdieu (1977) to refer to the different artifacts and approaches that people implement to achieve higher social status. This paper illustrates how cultural capital is an asset in the learning of English in Puerto Rico. The methodology included classroom observations, a sociolinguistic questionnaire, sociolinguistic interviews with focal parents, and brief interviews with focal children. The conclusions discuss how English is perceived as a social mobility tool in the social class strata of Puerto Rico.

Keywords: English in Puerto Rico, bilingualism, speech communities, cultural capital, social class

1. INTRODUCTION. This paper describes the educational, sociocultural, and linguistic factors that influence the learning of English as a second language in the public and private schools of Puerto Rico. The following research questions guided this research: (a) Are there two speech communities in Puerto Rico? If so, what are the language use and social differences in these two speech communities? (b) How do public and private school children and their families see English as a social mobility mechanism in their future professional lives? (c) What kind of interaction do public and private school children have in their English classrooms? This study presents my belief that the access to additional resources—English tutoring, English literacy, bilingual professional parents, traveling abroad, extracurricular activities—is fundamental in the development of private school students' linguistic, social, and educational skills.

2. LITERATURE REVIEW. Brutt-Griffler (2002) suggests that Puerto Rico is developing two speech communities via a macroacquisition process. A group sharing a particular language is a SPEECH COMMUNITY (Brutt-Griffler 2002:141). The speech community is a linguistic expression of the community. Brutt-Griffler defines MACROACQUISITION as the acquisition of a second language by a speech community; it is a second-language acquisition process. She illustrates two types of bilingual speech communities. Type A macroacquisition takes place in a multilingual setting. Type B refers to macroacquisition that takes place in a monolingual society, and then the acquisition of a second language turns that society into a bilingual speech community. Puerto Rico would be classified as a Type B bilingual speech community. Brutt-Griffler suggests that in macroacquisition contexts, the children of emerging speech communities acquire their English from schools. In many countries, schools might be the only English-speaking context that are emerging speech communities and are developing new English speakers. However, in Puerto Rico, as in other countries colonized by the United States or England (e.g. India, Hong Kong, Sri Lanka), effective English education is reserved for the wealthier classes, while the working classes receive limited English education (Brutt-Griffler 2002, Pousada 1996, 2009, Schweers & Hudders 2000, Torruellas 1990). Two proposed speech communities are used in this study: the

lower-working class families and their children (public school families) and the upper-middle class families and their children (private school families). The criteria used to cast a family as being of a particular class is discussed in 3.1.

Previous studies about English learning in Puerto Rico from the perspective of colonialism and Americanization do not make an in-depth analysis of many language questions, such as: (a) Is English a foreign or a second language on the island? (b) What is the relationship between the use of English and social class? (c) Is there any resistance or empathy toward English in schools? (Kerkhof 2001, Nickels 2005, Pousada 2009, Torruellas 1990, Zentella 1999).

Cultural capital refers to the different artifacts and approaches that people implement to achieve higher social status (Bourdieu 1977). New studies should critically examine English as a form of cultural capital in Puerto Rico and as a means of reproduction of social inequality (Torruellas 1990). Several authors expose the relationship between language choice and social class in Puerto Rico (Torruellas 1990, Resnick 1993, Pousada 1996, 2009, Ramírez-González and Torres-González 1996, Schweers & Hudders 2000). Segregation exists between those who have mastered English and those who have not. Highly competent bilinguals in Puerto Rican society tend to be upper-class members of the elite. Although English instruction is available in public schools as a mandatory subject, studies indicate that students from private schools more often become functional bilinguals because private schools teach English effectively. These private school students have access to forms of cultural capital – bilingual parents, private tutors, English literacy, traveling abroad, and extracurricular activities – and this increases their opportunities to practice their English (Torruellas 1990, Pousada 2009). Children of the elite continue to become bilingual and to be able to attend universities in the United States. The problem of learning English in Puerto Rico is not limited to merely a personal choice of whether to speak Spanish or English; it goes much further and deeper. Learning English in Puerto Rico involves complicated historical, social, cultural, economic, and educational factors that are intertwined with issues of social-class stratification.

This research contributes to the current literature related to English as a second language and to social-class stratification in Puerto Rico. This work is significant to the field of educational linguistics and sociolinguistics because only a few studies (Torruellas 1990, Pousada 1996, 2009) compare access to additional educational and social resources (e.g. bilingual parents, private tutors, English literacy, traveling abroad, and extracurricular activities) provided by public and private elementary-school students in Puerto Rico and its relationship with the learning of English.

3. METHODOLOGICAL APPROACH. I used qualitative research to conduct this study. I study the participants in their natural settings, which helped me build a holistic picture of their communities (Creswell 2007). This research uses an ethnographic approach that involves the study of a particular cultural group (Creswell 2007). I studied two social and economic groups that included three public school families and three private school families in rural Puerto Rico. I conducted the research in two small towns, Camuy and Hatillo, on the north coast of Puerto Rico. This research used the critical ethnography approach, which studies schools that provide certain privileges and addresses concerns of inequality, power, and dominance (Creswell 2007). I sought to determine if the private school provided certain privileges (learning English) to its students (upper-middle class children) that the public school did not.

3.1. CRITERIA FOR SELECTION OF THE FOCAL PARTICIPANTS. I selected three families in each of the schools because two families would only provide a small number of participants. After I obtained the approval of the Human Research Protections Office (HRPO/IRB) at the University of New Mexico, I started my research. To select the focal children and their families, I designed a brief questionnaire in which the parents answered questions about their children's education and English learning. During the first week of my observations in the schools, I gave 22 notification letters and brief questionnaires to all the children in the class to give to their parents. They returned the notification forms and questionnaires the next day of class. I analyzed the responses to the questionnaire. One of the main selection criteria was socioeconomic status. I did my best to select children who represented the population of each school. The school population demographics were determined with the help of the school administrators and English teachers. I consciously selected lower-working class participants at the public school and upper-middle class participants from the private school.

3.2. INSTRUMENTS AND DATA COLLECTION PROCESS. The three focal families at each school consented to be part of the research and to be interviewed. These focal children were learning English as a second language. All were fifth graders at Central Town Public School and Ocean View Private School.[1] I observed the English classroom twice a week for four months (September 2011 to December 2011). The duration of each observation was approximately one hour to one-and-a-half hours, depending on the school. I conducted 23 hours of observation in each school. I conducted two sociolinguistic interviews with one parent of each of the students selected as participants and two brief interviews with the six children at their homes. These were individual interviews with each one of the focal participants. I, a native speaker of Spanish, conducted all of the interviews in Spanish (the participants' native language). I carried out the first set of interviews in October 2011 and the follow-up interviews in February and March 2012. I focused on certain aspects of their social lives, such as their view of English, English as a social mobility tool, extracurricular activities, use of technology in their homes, and the use of Spanish and English. The first interviews included the same set of questions for every focal child and the same set of questions for every focal parent. The second interviews included a different set of questions for each focal student and a different set of questions for each focal mother and/or father based on the responses of the first interview questions. All of the parents and children's interview questions were open-ended. A sociolinguistic questionnaire was administered to the six focal parents before the first interview. The questionnaire included multiple-choice questions about demographic facts, parents' education, use of technology, leisure activities, the families' uses of English, Spanish, literacy, and educational practices.

4. THE PARTICIPANTS. The six focal families are the main source of information for this research. They are a sample of the families whose children attended Central Town Public School and Ocean View Private School at the time of the study. The details in the following two tables point out the main sociocultural characteristics of these families. As shown in Tables 1 and 2, each family had unique characteristics and a particular interest in English and in the education of their children.

Focal family	Number of family members	Parents' jobs	Access to cable TV/ Internet	Extra-curricular activities	Traveling/ vacation
Family of Joshua	Mother, father, two boys	Mother: receptionist Father: driver	Cable TV: yes Internet: yes	No	Yes – one time to Orlando, FL.
Family of Aidan	Mother, father, three boys	Mother: stay at home mother Father: mechanic	Cable TV: yes Internet: yes	Yes – Music classes funded by a federal program.	No
Family of Vincent	Mother, one boy, one girl	Mother: cashier	Cable TV: yes Internet: no	No	No – But they lived two or three years in California.

TABLE 1. Comparison of Public school Focal Families.

Focal family	Number of family members	Parents' jobs	Access to cable TV/ Internet	Extra-curricular activities	Traveling/ vacation
Family of Jessica	Mother, father, two girls, one boy	Mother: secretary Father: pensioned	Cable TV: yes Internet: yes	Yes – Dance classes, volleyball team, church plays	Yes – two times to Orlando, FL. Summer weeks in a vacation complex on the southwest coast of Puerto Rico
Family of Amanda	Mother, father, and two girls	Mother: teacher Father: director of a technology center	Cable TV: yes Internet: yes	Yes – Volleyball team, guitar classes	Yes – Orlando, FL and Dominican Republic.
Family of Jason	Mother, Father, and Three boys	Mother: stay at home mother Father: engineer	Cable TV: yes Internet: yes	Yes – Art classes, catechism	Yes – California and Dominican Republic.

TABLE 2. Comparison of Private school Focal Families.

The parents' jobs and the children's extracurricular activities were two of the main differences between the two social groups. The jobs of the parents in the private school required higher levels of education, expertise, and the use of English. Private school parents were more exposed to the use of English in their jobs (secretary, director of a technology center, and engineer) than the parents in the public school. All of the focal children in the private school attended art-related or sport-related extracurricular activities that their parents paid for. Only one public school child attended an extracurricular sport-related activity that was federally funded. Private school focal children had a more extensive interaction with children of their same or higher social status than the public school focal children. These private school children experienced more diverse academic and social contexts than the public school focal children. The diverse contexts bring an additional opportunity for private school focal children to practice English sporadically with different children, as stated by two of the parents.

5. CLASSROOMS OBSERVATIONS. To describe what kind of interaction public and private school children have in their English classrooms, this section presents descriptions of the children's social interactions inside their classrooms. This analysis presents the numbers of the participation of the students in English and Spanish. The numbers and percentages were calculated in the following manner: every time a student participated in the class and answered one of the teacher's questions (e.g. What is the setting of the story?) and/or completed grammar exercises (e.g. Fill in the blanks with the correct form of the verb "be") in Spanish or English, I tallied that response in a chart with a check mark. I created a new chart each day I observed a class, which was divided into columns for Spanish, English, and no response. I then tallied the number of responses for all 23 hours of observations over four months. Table 3 presents the totals for participation and student answers in Spanish and English for the students in the public and private school English classrooms.

Central Town Public School			Ocean View Private School		
Spanish	**English**	**No Response**	**Spanish**	**English**	**No Response**
155	**260**	11	93	**469**	0
36%	**61%**	3%	17%	**83%**	0%
Total number of responses = 426	**Total number of responses = 426**	Total number of responses = 426	Total number of responses = 562	**Total number of responses = 562**	Total number of responses = 562

TABLE 3. Observations in the English Classroom.

As the chart shows, the fifth grade students at Central Town Public School gave their answers most of the time in English. The students answered in English 61% of the time during those four months and in Spanish 36% of the time. They did not respond to the teacher's questions 3% of the time. The students at Ocean View Private School provided their answers primarily in English. The students answered in English 83% of the time during those four months and in

Spanish 17% of the time. The students at Ocean View Private School always responded to the teacher's questions.

5.1. OTHER CLASSROOM OBSERVATIONS AND FINDINGS. The observations in both English classrooms helped to distinguish the two speech communities. The interactions and patterns in the English classrooms of both schools are different and diverse. The main differences are reflected in the students' uses of English inside the classroom, the students' reactions to the participation of their classmates, and the students' opinions and actions with respect to the English class teaching methods.

Central Town Public School English Classroom	**Ocean View Private School English Classroom**
Social interaction – Students did not move around the classroom a lot; no cooperative learning; students felt uncomfortable and sometimes were laughed at by classmates when speaking English; frequent interaction in English and Spanish.	***Social interaction*** – Students moved freely in the classroom; frequent cooperative learning; students helped each other when someone did not know the correct pronunciation of a word; most of the speaking interaction was in English.
Participation – Sometimes students did not answer the teacher's questions; infrequent use of spoken English; Vincent was the focal student who used English the most.	***Participation*** – Students were excited to participate; more spoken English; Jason was the focal student who used English the most.

TABLE 4. Social Interactions in the Two English Classrooms.

One main difference between the classrooms was the role of the students in the learning-teaching process. The students in the English classroom in the public school did not seem to be very active in terms of participating and speaking in English. Cooperative learning or teamwork was almost absent in this classroom, and students who participated individually were sometimes laughed at and made to feel uncomfortable while trying to practice and improve their English. In contrast, the students in the English classroom of the private school acted as active agents in their learning of English as a second language. These students interacted more with their classmates and were able to move from one place to another inside the classroom to check and discuss grammar and literacy tasks with their classmates.

The private school students also acted as active agents in their learning when they frequently asked for the meanings of new words in English and when they encouraged the participation of their classmates. In contrast, the students in the public school checked their answers only with the teacher. Even though the public school students expressed their interest in the meanings of new vocabulary words, there was little encouragement for them to go beyond and participate more in the English class.

The participation of the focal students in Central Town Public School primarily involved the routine of selecting the correct word or phrase as an answer with little time for students to generate complex sentences as correct answers. Vincent's participation in the English class was more evident because he was more active reading aloud, and sometimes his classmates laughed

at him when his pronunciation was not correct. The focal students behaved well in the English classroom and followed the patterns and rules expected of them.

The participation of the focal students in Ocean View Private School was more noticeable because of the constant interaction and discussion of the English tasks among students. Jason was one of the most popular students in the English classroom because of his academic excellence. His classmates knew about Jason's highly proficient skills in English, and they always wanted to double-check their answers with him before the teacher corrected them.

6. PUBLIC SCHOOL FAMILIES. The perception of bilingualism of the public school focal parents included speaking English to be able to communicate and to find a good job. Public school parents did not mention during our interviews that their children speak English in their everyday context. These parents did not mention the use of English literacy (books, writing papers, technology) at the college level.

The parents in the public school stated reasons for considering English to be an important language in their children's education. These parents focused mainly on the need for English for communicative purposes in the workplace. They also considered English to be an important language in some national and federal jobs. Parents in the public school mentioned English as an important tool if their children move to the United States. The parents offered these answers to some of the interviewer's open-ended questions: (a) Do you support your children learning English in school? Why? (b) Why do you think learning English is important for their future life? (c) Do you think English is important to find a better job with higher pay? Why?

Parents' answers:
- Jobs and professions: education jobs, government jobs, federal jobs, U.S. Postal Service, Army.
- To move to another country.
- To work in the United States.
- To practice English in the United States.
- To read documents in English.
- To speak with someone from the United States in the workplace.

The public school focal children described their views of English in terms of socialization and academic purposes. The focal children also offered different answers to some of the interviewer's open-ended questions: (a) Do you think English is important for you? Why? (b) In what job positions/fields do people use English more? Why?

Students' answers:
- To speak with someone who speaks English in the workplace.
- To move and live in the United States.
- To play professional sports in the United States.
- To work in professional fields such as law and in the Army.

Parents as well as children focused on jobs in the military and postal service. The public school families said that the main goal is to speak very good English to communicate with any English speakers in Puerto Rico or in the United States. For these families, speaking English is

seen as a golden ticket that can open doors for any job position, especially in federal government branches.

7. PRIVATE SCHOOL FAMILIES. Private school focal parents mentioned bilingualism and the components of a successful bilingual education, such as: reading, speaking, vocabulary, writing, good pronunciation, and grammar. Some of these parents and their family members classify themselves as bilingual and feel comfortable speaking English. Two of the parents reported that their children speak English and most of the time felt comfortable using the language. According to these families, speaking is an important part of being bilingual. These focal parents are aware of other skills that are important to becoming a fully bilingual person, such as correct pronunciation, vocabulary, and literacy. These focal parents connected the use of English literacy with the future of their children at a university. They said how English literacy would be an advantage in their children's college life.

The focal parents in the private school gave diverse reasons for considering English a fundamental language in their children's future. Parents were focused on the use of English at the college level and as an important language in high paying professional careers. Finally, two of the focal parents said knowing English is the first step to becoming a multilingual speaker and having access to cultural knowledge from around the world. The parents offered these answers to some of the open-ended questions during the interviews: (a) Do you support your children learning English in school? Why? (b) Why do you think learning English is important for them? (c) Do you think English is important to find a better job with higher pay? Why?

Parents' answers:
- To move to the United States.
- To find a good job.
- To attend a good college.
- To have an easier life during the college years.
- To take the Advanced Placement Exams and transfer college credits in English courses.
- To have access to material possessions.
- To enter the fields of medicine, engineering, law, accounting, business, education, architecture, pharmacy, and to obtain a master's degree and a Ph.D.
- To travel, experience different cultures, meet new people, be accepted into internships, to study abroad.
- To understand technology and media.
- To work in multinational companies.
- To learn other languages besides English, such as Japanese, Portuguese, Mandarin, and French.
- To recognize bilingualism as a tool for career advancement.

During their interviews, the focal children in the private school expressed their thoughts about English at the university level and the need of English to become a multilingual person. The focal children also answered the interviewer's open-ended questions: (a) Do you think English is important for you? Why? (b) In what job positions / fields do people use English more? Why?

Students' answers:
- To use it in the university.
- To study in a university in the United States.
- To use it in the following professional fields: medicine, law, business, technology, and computer programming.
- To travel and to experience other cultures.
- To be multilingual.

The perspectives about English of the private school children are similar to those of their parents. Parents as well as children see the need for English in their future academic lives in college and in high paying professional jobs.

8. CONCLUSION. To conclude, it is necessary to look back to the research questions to answer them with the analysis of the data provided. The first research questions: Are there two speech communities in Puerto Rico? What are the language use and social differences in these two speech communities? The following table presents the main findings of the two proposed speech communities:

Lower-Working Class/Public School	**Upper-Middle Class/Private School**
Pre-existing monolingual Spanish speech community.	Type B bilingual speech community.
Limited access to extracurricular activities.	Extracurricular activities.
Parents are not bilingual.	Parents are Spanish-English bilinguals.
Limited interaction in English in the classroom.	Cooperative learning and frequent English interaction in the classroom.
No clear connection of English literacy with college education.	English literacy important for college education.

TABLE 5. Proposed Speech Communities.

Brutt-Griffler (2002) makes clear that she is not calling countries such as Mexico, Japan, and Jordan bilingual nations. She is pointing out the existence of bilingual speech communities within them. Puerto Rico has a bilingual speech community within the monolingual Spanish speech community. Brutt-Griffler (2002:147) adds that 'bilingual' in the speech communities context does not refer to 'balanced bilinguals' who possess equal fluency in each language. According to my observations and the data from the interviews, the upper-middle class community is a Type B bilingual speech community under Brutt-Griffler's conditions. They are an existing bilingual community that has learned English as a second language in Puerto Rico. They are non-native speakers of English, and perhaps some of their English teachers are or were non-native speakers of English. The lower-working class is a separate community and does not fit into any of Brutt-Griffler's speech community categories. Many of them are Spanish monolinguals and do not share both languages – Spanish and English – with other members of the same community. It seems as if the lower-working class is the existing monolingual Spanish

linguistic community. The lower-working class is acquiring English as a second language, but they are acquiring it at slower pace and with a limited proficiency level when compared to the upper-middle class speech community.

The second research question was how do public and private school children and their families see English as a social mobility mechanism in their future professional lives. The public school focal families are an example of a speech community that sees English as a social mobility mechanism that can guarantee a secure job in a federal branch (postal service, Army). These families see English as a language that can bring economic stability to their children's lives, and this is also a way to see English as a social mobility mechanism to higher social classes, although not the elite social classes. The private school focal families are an example of a speech community that sees English as a door opener for many great opportunities in their children's lives such as high-paying professional jobs, a fulfilling college experience, and the opportunity to learn more languages. English is seen as a social mobility mechanism from upper-middle class to the upper and elite social classes of Puerto Rico.

The third research question asked what kind of interaction do public and private school children have in their English classrooms. The data from the public school English classroom revealed that students did not feel comfortable speaking in English because sometimes their classmates made fun of them. Also, conversation in Spanish and English was a common pattern. In contrast, the students in the English classroom of the private school seemed to have fun while learning and speaking their second language. The teamwork and cooperation in the English classroom of the private school were essential in the English learning process.

8.1. LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH. Some of the limitations of this study included the small number of families selected to participate. This is a qualitative research study and does not attempt to generalize the results of the sociocultural practices of the focal families to all of the lower-working class and the upper-middle class families in Puerto Rico. This research covered only two small towns in Puerto Rico, so the additional educational and social resources of the families and schools can be different depending on the geographic location and access to resources in bigger cities. Future research includes expanding the study to other geographical locations in Puerto Rico, including metropolitan areas and elite schools where access to English as a form of cultural capital is extensive and diverse.

REFERENCES

BOURDIEU, PIERRE. 1977. Cultural reproduction and social reproduction. *Power and ideology in education,* ed. by J. Karabel and A. H. Halsey, 487–511. New York: Oxford University Press.

BRUTT–GRIFFLER, JANINA. 2002. *World English: A study of its development*. Clevedon, UK: Multilingual Matters.

CRESWELL, JOHN. 2007. *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.)*.* Thousand Oaks, CA: Sage Publications.

KERKHOF, ERMA. 2001. The myth of the dumb Puerto Rican: Circular migration and language struggle in Puerto Rico. *New West Indian Guide* 75.257–288.

NICKELS, EDELMIRA. L. 2005. English in Puerto Rico. *World Englishes* 24.227–237.

POUSADA, ALICIA. 2009. *Dónde están los bilingües? La cartografía del idioma en Puerto Rico*. Ponencia presentada en la 8va Conferencia de la Asociación de Estudios Puertorriqueños, San Juan.

POUSADA, ALICIA. 1996. Puerto Rico: On the horns of a language planning dilemma. *TESOL Quarterly* 30.499–510.

RAMÍREZ–GONZÁLEZ CARLOS M., and ROAMÉ TORRES-GONZÁLEZ. 1996. English under U. S. sovereignty: Ninety-five years of change of the status of English in Puerto Rico. *Post- Imperial English: Status change in former British and American colonies, 1940–1990,* ed. by J. Fishman, A. Conrad and A. Rubal-López, 173–204. New York: Mouton de Gruyter.

RESNICK, MELVIN. 1993. ESL and language planning in Puerto Rico. *TESOL Quarterly* 27.259–73.

SCHWEERS, WILLIAM C., and MADELEINE HUDDERS. 2000. The reformation of democratization of English education in Puerto Rico. *International Journal of the Sociology of Language* 142.63–68.

TORRUELLAS, ROSA M. 1990. *Learning English in three private schools in Puerto Rico*. New York, NY: New York University dissertation.

ZENTELLA, ANA. C. 1999. Language policy/planning and US colonialism: The Puerto Rican thorn in English only's side. *Sociopolitical Perspectives on Language Policy and Planning in the USA,* ed. by T. Huebner and K. Davis, 155–171. Philadelphia: John Benjamins

[1] I used pseudonyms for the schools' and focal participants' names.

Analyzing prefabricated language in local advertisements

Karol Ibarra Zetter and Anni Leming
University of New Mexico

ABSTRACT. The present study investigated prefabricated language in the 'I saw you' section, in the Weekly Alibi, a magazine based in Albuquerque, New Mexico. The 'I saw you' section is comparable to missed connections that are commonly found in the romantic personal advertisements of newspapers. Formulaic language, also called prefabricated language entails a 'sequence of words that are stored and retrieved from whole memory and are a dynamic response to the demands of language use' (Wray & Perkins 2000:5). We were interested in the social functions, prefab types, and distribution of prefabs in the context of romantic personal ads. We found that distribution of prefab types in our data was akin to that found in Erman & Warren 2000, where the lexical type was most frequent. We compared our samples of most frequent constructions to different corpora and found similar patterns in one of the corpora.

Keywords: Formulaic language, prefabricated language, personals advertisements, taboo topics

1. INTRODUCTION. This study examines the use of FORMULAIC LANGUAGE, also known as PREFABRICATED LANGUAGE, in local personal advertisements. Literature on the subject has noted the HOLISTIC PROCESSING nature of formulaic language. Sequences of words are processed as units rather than parsed (analytically) due to their frequency of use, thus reducing cognitive load. Selecting a READY-MADE set of phrases and words that many other people use in order to communicate is a safe way of communicating with someone you do not know – yet. It is not surprising, then, that 55–70% of communicators will depend on these ready-made set of phrases – as show by numerous studies (e.g. Erman & Warren 2000, Wray & Perkins 2000) and in various contexts. Other scholars have found 20–40% formulaicity in unplanned native speech and academic writing (Foster 2001, Biber et al. 1999, Howarth 1998).

Wray and Perkins (2000:13) suggest that, 'we use [formulaic language] to frame [ideas] in a way that will maximize the likelihood of the required event coming about.' Simply stated, formulaic language facilitates the way we convey our messages depending on particular contexts. Additionally, there are cognitive processing advantages in the use of prefabricated sequences over creative language. There are SOCIO-AFFECTIVE advantages as well. That is, there are social and emotional consequences of language use, which are detailed below by Wray and Perkins (2000).

Wray and Perkins (2000:13–14) suggest that we use language to meet physical needs as well as to assert individuality and group belonging so as to 'ensure that we neither become subsumed within, nor are excluded from, the social networks which we feed off emotionally, and which directly contribute to our success in finding a reproductive partner.' Finding a partner is an important feature of our study. In other words, 'I saw you' advertisements are a way to reconnect and communicate with potential partners. The study here investigates the distribution of prefab types, and discusses the socio-affective use of formulaic language in these local personal advertisements.

2. THEORETICAL BACKGROUND. A great amount of languages' expressions is constructed through prefabricated phrases or idioms. Bybee (2006:713) points out that 'both written and spoken discourse are characterized by the high use of conventionalized word sequences, which include sequences that we might call formulaic language and idioms, but also conventionalized collocations.' She further suggests that idioms and prefabs are stored as whole units in the brain. Idioms and prefabs may have metaphorical and literal meaning, respectively (e.g. *the bottom line, the big picture, a piece of cake*); that is, it is possible to keep using them in new contexts of use with expanded or shifted meanings depending on the experience the individual user has with the prefabricated sequence.

The work of Bybee 2006 reminds us of the frequency and routinization of prefabs. 'Prefabs are word sequences that are conventionalized, but predictable in other ways, for example, word sequences like *prominent role*, *mixed message*, *beyond repair*, and *to need help*' (Bybee 2006:713). Phrasal verbs and verb-preposition pairings, which 'are pervasive in English as well as in other languages, can be considered prefabs, though in some cases their semantic predictability could be called into question' (713).

Erman and Warren (2000) investigated the prominence that prefabricated language has in the production and structure of the text, spoken or written. They examined the distribution of prefabs in texts and categorized them according to structure and function. The authors assumed that we use two resources to produce language: the open choice (word-to-word combinations) and preconstructed phrases. Erman and Warren (2000:31) define a prefab as 'a combination of a least two words favored by native speakers in a preference to an alternative combination which could have been equivalent had there been no conventionalization.' They distinguish different type of prefabs: LEXICAL, GRAMMATICAL, PRAGMATICS and REDUCIBLES. They contend that the exposure to prefabs is probabilistic. Therefore, identifying prefabs is a complicated task because not all the speakers are exposed to the same type of language.

Wray and Perkins (2000) assert that formulaic language is a tool of social interaction. Formulaic language entails word strings that possibly are registered in memory as a unit. Wray and Perkins' (2000:1) definition of formulaic language is a 'sequence continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language of grammar.' The authors noted that some of the formulaic language is more transparent than others. Besides, 'possibly as much as 70% of adult native language may be formulaic' (Altenberg 1990 cited in Wray & Perkins 2000:2).

Wray and Perkins (2000) claim that there is a continuum of formulaicity in functional expressions, composite units, lexical collocations as well as grammatical collocations. These authors describe the following social functions for prefabricated language.

(1) Manipulation of others: the effect of language to satisfy and emotional or physical need
(2) Asserting separate identity: the effect of language to be taken seriously
(3) Asserting group identity; the effect of language to show one's membership in a group (14).

According to Wray and Perkins (2000), the purpose of formulaic language is to help with communication and to avoid processing overload.

Among the studies of taboo language (Anderson & Trudgill 1990, Freitas 2008, Fleming & Lempert 2011) a commonality is to describe the use of certain mitigating strategies in order to convey a message without facing an explicit social transgression. The objective and function might be to fulfill point (2) above in Wray & Perkins's 2000 model. Certain kinds of speech have been regulated by different institutions, such as state, religion, social conventions, etiquette, etc. (Fleming & Lempert 2011). However, as Fleming and Lempert (2011:6) mention,

> Ironically, proscriptions and even the appropriate substitutes these regimes recommend (e.g. euphemisms, circumlocutions, special citational forms like 'the F-word') make taboo utterances more salient. And rather than fix or stabilize a speaker's relation to the taboo object—by ensuring a safe, respectful 'distance,' for instance—proscription and efforts at containment seem to make such relations less stable. As conventions, they may now be flouted, parodied, played upon, or otherwise altered for strategic and interactional effect.

Fleming and Lempert (2011:6) remark that pragmatic prefabs or ready-mades have 'context coiled tight inside' for which producing those expressions in different contexts could be considered as a transgression of their primary social function, e.g. declaring a couple married by a non-minister; cursing in public broadcasts; pronouncing the quotation of a phrase that includes cursing, thus reduplicating the cursing and making the taboo word more outstanding (6–8).

Freitas (2008) looks at various modes in advertising to analyze the influential role of taboos. Freitas offers a definition of taboo as being 'a number of restrictions that regulate some areas of social life, and that demand avoidance behavior and that can be a threat to the established social patterns' (39). Drawing on this definition, two types of taboo are explained. One is when something should be avoided because it is negative, whereas the other type is when something is still a social violation but enticing to people. Especially in personal advertisements of the Weekly Alibi, it was observed that speech conveyed taboos. In the following examples the people who wrote the advertisements made explicit references to what could be considered taboo (e.g. sex).

(1)   … Wanted to say hi but you were **handling the meat**, email me if you want to handle mine
(2)   You are stunning and your eyes paralyze me. Your energy cuts through me. **Let's get sweaty**?
(3)   I'll hoist my Jolly Roger. Now **let me have mah booty**!

However, these other advertisements downplayed a social taboo with the use of some prefabricated phrases (e.g. suggesting or inviting to have a more intimate relationship).

(4)   I know you're crazy fun so **let's get into some trouble** together.
(5)   We have lots of paddles, and straps, etc. to use. Please send a note with your interests and ideas and/or questions. **Experience isn't necessary,** only enthusiasm is.
(6)   I could tell by the stickers on your Mac Book that you like to snowboard. **Let's go for a ride** cowgirl.

For our particular study, we investigate the use of prefabricated language as a means of discussing taboo topics that are appealing rather than negative, as presented in the 'I saw you' advertisements, especially given the overall purpose of the advertisements. Our study examines the use of prefabs as a mitigating discourse strategy when speakers are approaching sensitive or taboo subjects in the social advertisements of a local magazine.

3. DATA AND METHODS. The purpose of this study is to investigate the use of formulaic language in the context of local 'I saw you' advertisements, personal ads about missed connections. The research attempted to answer three research questions:

> (1) What types of prefabs are present in 'I saw you' ads?
> (2) What is the distribution of these prefab types?
> (3) How are prefabs used to communicate sexual taboo topics?

3.1. DATA. The data were obtained from a local magazine called the Weekly Alibi. It is printed every week and it is freely available in different businesses and in the University of New Mexico area. The Weekly Alibi is also available online. The Weekly Alibi has different sections and one includes the Personal Ads for Dating, Friendship and 'I saw you', which is the missed connections section of the advertisements.

3.2. METHOD. A database of 100 'I saw you' advertisements from the Albuquerque Weekly Alibi was compiled. We randomly selected at least two issues per month from 2009 until 2012 in paper-based form or online, and from each issue we randomly selected a maximum of four ads.

Each advertisement was analysed by the researchers for the presence of prefabs. In the selection process a co-referentiality check was done to ensure both researchers agreed on each prefab. A total of 148 prefabs were extracted from the advertisements, along with one to two sentences of the passages in which they were found to provide context.

The analysis was based on the prefab categories detailed below by Erman and Warren (2000). We calculated the use of prefabs by dividing the number of each type by the total number of 148 prefabs (see Table 2).

The coding categories in our analysis included: number of prefabs in the advertisement, type of prefabs (Erman & Warren 2000), and social functions of the prefabs (Wray & Perkins 2000).

The types of prefabs that we utilized were LEXICAL, GRAMMATICAL, and PRAGMATIC. Erman and Warren (2000:38–43) offer the following definitions of prefab types:

> (1) LEXICAL PREFABS: 'semantic units that … denote entities, properties, states, events, and situations of different kinds.'
> (2) GRAMMATICAL PREFABS: 'intralinguistic text-forming items rather than units with extralinguistic reference.'
> (3) PRAGMATIC PREFABS: 'functional in that they do not directly partake in the propositional prefabs in that they may occur outside the syntactic structure.'

Table 1 provides an example of Erman and Warren's (2000) examples of these prefab types and examples of these types in our data.

Type of prefab	Erman & Warren (2000)	'I saw you' Ads
LEXICAL	Out of date, run off, permanent job, at the time	Out of practice, you must be taken, be mine
GRAMMATICAL	A few, a great deal, for instance, be going to, sort of, let us	Let me [have mah booty], had to run
PRAGMATIC	I'm afraid (to tell __), yeah, quite, right	I must say well, of course
REDUCIBLES	I'm, It's, Don't, They've, let's	You're, I'm, I'd, let's

TABLE 1. Types of prefabs.

After determining that the lexical prefab was the most frequent, we chose the five most frequent prefabricated sequences within our data and conducted a corpus analysis to determine if the distribution of our five most frequent would match the distribution found in different corpora. The corpora utilized were the Corpus of Contemporary American English (COCA) (Davies 2008-), the British National Corpus (BNC) (Davies 2004-), and the Michigan Corpus of Academic Spoken English (MICASE) (Simpson et al. 2002).

The following are the most frequent prefab constructions: let/let us, take + NP, make/made me/you + V, would love to + VP, and couldn't + VP. For token constructions that varied in type use, e.g. *take me, take my*, each type was counted and divided by the total number of type use to get an average.

4. RESULTS. We found that the prefab types found in the Weekly Alibi pattern with the same distribution found in Erman & Warren 2000 (see Table 2). Lexical prefabs were used the most, grammatical prefabs the second most, and the least used were pragmatic prefabs.

Because lexical prefabs were used the most, we decided to mine for the five most frequent prefabs in our data and compare their constructions in the three major corpora previously mentioned (see results in Table 3). In the case of COCA comparison, the *let* construction was the second most frequent and *couldn't* the least.

In the instance of the MICASE comparison, the *would love to* construction was the least frequently used prefab and the *couldn't* construction was the second least frequent. The BNC corpus then seems to pattern the same frequency distribution of our most frequent constructions.

Use of prefabs according to Erman and Warren (2000)				Alibi Data	
	Speech	Written text	Total	Written text	Total
LEXICAL	38.8%	71.5%	71.5%	127/148	85.8%
GRAMMATICAL	16.9%	20.5%	16.9%	18/148	12.2%
PRAGMATIC	2.4%	16.7%	Together	3/148	2%
REDUCIBLES	9.2%	24%	2.4%	N/A	N/A

TABLE 2. Distribution of prefab types.

	Let +me/us	Take +NP	Make	Would love to +VP	Couldn't +VP
Most Frequent Prefab/ Total Prefabs	**7.4%** (11/148)	**6.7%** (10/148)	**6.1%** (9/148)	**4.7%** (7/148)	**3.3%** (5/148)
Most Frequent Prefab/ Total Most Frequent	**26.1%** (11/42)	**23.8%** (10/42)	**21.4%** (9/42)	**16.7%** (7/42)	**11.9%** (5/42)
Most Frequent Prefab/ COCA 400 M (million words)	**.000019%** 773/400 M	**.0000069%** 275/400 M	**.0000053%** 214/400 M	**.0000027%** 109/400 M	**.000022%** 897/400 M
Most Frequent Prefab/ BNC 100 M (million words)	**.00029%** 2945/100 M	**.000079%** 795/100 M	**.000053%** 534/100 M	**.000023%** 237/100 M	**.0000099%** 99/100 M
Most Frequent Prefab/ MICASE 1.8 M (million words)	**.0080%** 1454/1.8 M	**.0078%** 1411/1.8 M	**.0069%** 1257/1.8 M	**.00005%** 9/1.8 M	**.0016%** 292/1.8 M

TABLE 3. Analysis of the five most frequent prefabs in Alibi data.

The following sentences are provided in order to gain a detailed sense of the use of our five most frequent prefabs.

- I'll hoist my Jolly Roger. Now ***let me*** *have mah booty*!
- Your amazing Bruce Lee physique ***made my*** *head spin.*
- Watched you biking down the street, ***would love to*** *get to know you ....*
- I winked at you at the stop light. Wanna ***take me for a ride*** sometime?
- I was there with my primo and he was laughing cuz I ***couldn't take my eyes off you***.

5. DISCUSSION. The analysis above focuses on the prefab types and distribution of them; however, we must not forget to consider their function in communication. Advertisements such as those used in this study require participants to produce written communication efficiently in a small space in the publication, much in the same way speakers rely on formulaicity in conversations, which tend to have more time constraints than in formal writing. Furthermore, it is interesting that our five most frequent prefabricated sequences are also among the five most frequent in a major corpus, indicating that prefabricated language provides an avenue for conveying taboo messages to their imagined addressee. It appears that prefabs 'facilitate production and presumably the interpretation of utterances' (Erman & Warren 2000:50).

5.1. STRUCTURE AND DISTRIBUTION. Drawing on the work of Erman & Warren 2000, the genre of 'I saw you' written ads indicates that there is a pattern of distribution. The following pattern of distribution was found in the prefabs of the Weekly Alibi: first, the lexical prefabs, then grammatical, and last the pragmatic prefabs. Recall that lexical prefabs are 'semantic units that

… denote entities, properties, states, events, and situations of different kinds' (Erman & Warren 2000:38). In the context of the 'I saw you' advertisements, it makes sense that advertisers rely on language that refers to events or different situations in which they met or saw their intended addressee.

5.2. Socio-affective. The findings in this small-scale study indicate that the prefabricated language offers a way to mitigate the seemingly taboo nature of personal advertisements, which may be threatening to the people who wrote the advertisements. Freitas (2008:39) defined taboo as 'a number of restrictions that regulate some areas of social life, and that demand avoidance behavior and that can be a threat to the established social patterns.' As an alternative to using prefabricated language as an avoidance behavior, its use in an advertisement may tend to have stronger implications. Fleming and Lempert (2011) note that although social proscriptions to mitigate taboo language are meant to provide 'safe, respectful "distance"' to the object referred to, the effects may have a stronger force, especially when considering the intention of personal advertisements.

Regardless of whether prefabricated language is used to avoid directly mentioning or strongly implying one's desire for a potential partner, it is used as a means for physical and social survival, especially when managing interactions with a potential partner through the written advertisement. Referring back to Wray and Perkins (2000:14), who described the social functions for prefabricated language, it appears that the major function is to use 'language to satisfy an emotional and physical need.'

6. Conclusions. Social interaction can be performed utilizing formulaic language. The Weekly Alibi's personal advertisements reflected the social function already discovered by some studies on formulaic language and prefabs. The personal advertisements also mirrored the distribution of Erman and Warren's (2000) categories of prefabs. Prefabs are not only a way to reduce the working memory and cognitive load in speech; they are also a means to avoid emotional load.

According to Wray and Perkins (2000:15), adults have to frame their requests, petitions, etc. in a way that is more subtle than a simple and straightforward claim, due to 'the hierarchical relationships that exist between communicating adults.' In this sense, some of the people that post their advertisements of missed connections, using prefabs – familiar and immediately accessible language – could be softening their wishes to meet their physical or psychological needs.

The driving force behind the socio-interactional formulas is ensuring that the speaker gets what he/she wants and is perceived as an individual within the group (Wray & Perkins 2000:18). In a sense, prefabs are blocks of language that encompass metalinguistic functions. Prefabs help speakers to distance themselves with messages uttered through the conventionalization of the formulaic language. They also help social interaction through the conveyance of an individual or group identity, creating frameworks of social interpretation to maximize the likelihood of conveying the message, and being understood to avoid problematic interaction.

7. Directions for future research. Socio-affective use of prefabricated language in personal advertisements deserves more study. Considering the socio-affective and socio-cultural approaches to interpreting language may contribute to the discussion on prefabricated language. Factors such as gender, for example, may very well affect the way prefabricated language is used as a distancing mechanism with such taboo topics that are inherent in personal advertisements.

Next, rather than simply relying on the researchers for interpretation, survey participants might be invited to do perception tasks on the effect of prefabricated language. Furthermore, the present study is small-scale and exploratory in nature and would benefit from an increase in the number of advertisements and prefabricated sequences for analysis. These considerations would avoid the shortcomings of two researchers' intuitions and we would also gain a more accurate sense of how language functions socially.

In the present study, we analysed the use of prefabricated language and its varying combinations. Another consideration is idiomatic formulaic sequences, which tend to be processed holistically due their inherent characteristics and their meanings that 'cannot be derived from the sum of the individual words' (Conklin & Schmitt 2008:80) in sequence. Bybee (2006) states that although idioms are metaphoric in nature, there is a fine line between idioms and prefabs, indicating that both types of expressions are mentally stored. Another idea for future research, then, might address the distinct use of idiomatic prefabs versus those that have more decomposability in personal advertisements.

REFERENCES

ALTENBERG, BENGT. 1990. Speech as linear composition. *Proceedings from the Fourth Nordic Conference for English Studies (vol. 1),* ed. by G. Caie, K. Haastrup, A. L. Jakobsen, J.E. Nielsen, J. Sevaldsen, H. Specht, and A. Zettersten, 133–143. Copenhagen: Department of English, University of Copenhagen.

ANDERSON, LARS-GUNNAR, and PETER TRUDGILL. 1990. *Bad language*. Oxford, UK and Cambridge, MA: Basil Blackwell.

BIBER, DOUGLAS; STIG JOHANSSON; GEOFFREY LEECH; SUSAN CONRAD; and EDWARD FINEGAN. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

BYBEE, JOAN. 2006. From usage to grammar: The mind's response to repetition. *Language* 82.711–33.

CONKLIN, KATHY, and NORBERT SCHMITT. 2008. Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics* 29.72–89.

DAVIES, MARK. 2004-. *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). Online: http://corpus.byu.edu/bnc/.

DAVIES, MARK. 2008-. *The Corpus of Contemporary American English: 450 million words, 1990-present*. Online: http://corpus.byu.edu/coca/.

ERMAN, BRITT, and BEATRICE WARREN. 2000. The idiom principles and the open choice principle. *Text* 20.29–62.

FLEMING, LUKE, and MICHAEL LEMPERT. 2011. Introduction: Beyond Bad Words. *Anthropological Quarterly* 84.5–13.

FOSTER, PAULINE. 2001. Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing*, ed. by Peter Martín Bygate and Merrill Swain Skehan, 75–93. Harlow: Longman.

FREITAS, ELSA-SIMÕES-LUCAS. 2008. *Taboo in Advertising*. Amsterdam: John Benjamins.

HOWARTH, PETER. 1998. The phraseology of learners' academic writing. *Phraseology: Theory, Analysis and Applications*, ed. by Anthony Cowie, 161–86. Oxford: Oxford University Press.

SIMPSON, RITA. C.; SARAH. L. BRIGGS; JANINE OVENS; and JOHN. M. SWALES. 2002. *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan. Online: http://micase.elicorpora.info/.

WRAY, ALISON and MICHAEL R. PERKINS. 2000. The functions of formulaic language: an integrated model. *Language and Communication* 20.1–28.

A radial category profiling analysis of North Sámi ambipositions

Laura A. Janda, Lene Antonsen, Berit Anne Bals Baal
University of Tromsø

ABSTRACT. We present a study of four North Sámi adpositions that can be used as both prepositions and postpositions and thus be termed 'ambipositions'. We advance three hypotheses concerning 1) dialectal differences in use of ambipositions in North Sámi, 2) differences between their use as prepositions and postpositions, and 3) a possible typological correlation between the frequency of ambipositions and the extent to which position is used to differentiate meaning, with North Sámi at the high end of this scale. Our study tests these hypotheses against two databases representing the use of ambipositions in newspapers and in literature.

1. INTRODUCTION: NORTH SÁMI AND ITS AMBIPOSITIONS. This study examines the behavior of North Sámi adpositions with particular emphasis on the relationship between prepositional vs. postpositional use and the expression of meaning. The major tool used in this investigation is radial category profiling (Nesset et al. 2011, Janda et al. 2013), which emerges from the radial category model of meaning familiar to cognitive linguists (Taylor 2003, Langacker 2008).

North Sámi is an indigenous Finno-Ugric language spoken by approximately 30,000 people in contiguous regions of northern Norway, Sweden, and Finland. North Sámi is unique in Europe as a minority language that is in contact with majority languages from two different language families: Indo-European (Norwegian and Swedish) and Finno-Ugric (Finnish; Ylikoski 2009:201–202). This situation is significant for our study because the languages that North Sámi is in contact with exert opposite pressures on the use of adpositions: Norwegian and Swedish have primarily prepositions, whereas Finnish has primarily postpositions.

Ambipositions are adpositions that can appear as both prepositions and as postpositions. An English example is *over* which is a preposition in *He travelled all over the world*, but a postposition in *He travelled the world over*. While it is not unusual for a language to have some ambipositions, it is unusual for a language to make extensive and systematic use of ambipositions (Hagège 2010:116–124). Typically one of the positions is highly marginal, as we see with the postpositional use of English *over*. Russian, for example, shows somewhat more variation than English with over 150 prepositions like *na* 'on(to)' and *pod* 'under', one postposition *nazad* 'ago', and three ambipositions *spustja* 'after', *pogodja* 'after', and *radi* 'for the sake of'. However, both *spustja* and *pogodja* are marginal relative to the synonymous preposition *posle* 'after', and *radi* 'for the sake of' is mainly used as a preposition. In other words, prepositions are the norm in Russian, where postpositions are few and usually marginal. By contrast, the majority of Finno-Ugric languages make exclusive or nearly exclusive use of postpositions (Grünthal 2003:45).

Finnish, Estonian, and the Sámi languages stand out typologically because they make extensive use of ambipositions. While all three are predominantly postpositional languages, prepositions and ambipositions are used systematically. Finnish and Estonian have relatively similar distributions: Finnish has 76% postpositions, 10% prepositions, and 13% ambipositions while Estonian has 74% postpositions, 16% prepositions, and 10% ambipositions. The percentage of postpositions in North Sámi is similar at 75%, but the remaining proportions are

very different, with only 3% prepositions but 22% ambipositions (data compiled from Karlsson 2008:313–320, Grünthal 2003:57, Nickel & Sammallahti 2011:171–196). In other words, North Sámi makes much more extensive use of ambipositions than either Finnish or Estonian.

In both Finnish and Estonian the position of ambipositions tends to be correlated with expression of different meanings (Huumo 2013, Lehismets 2011, Erelt 2003:117–118, Grünthal 2003). For example, if an ambiposition can express both time and space, the tendency is to use the preposition to express time and the postposition to express space, as we see with the Finnish ambiposition *läpi* 'through' in *metsän läpi* [forest.GEN through] 'through the forest' vs. *läpi talven* [through winter.GEN] 'through the winter' (note that GEN = genitive). Very little has been written about ambipositions in North Sámi; exceptions are Ylikoski 2006, Nielsen 1979, and Bartens 1974, but aside from the comment that prepositional use can be more emphatic than postpositional use (Nielsen 1979:188–189) none of these works address differences between prepositional and postpositional use in any detail. Ours is the first study to focus on the relationship between position and meaning in North Sámi ambipositions.

Given what is known about the contact situation of North Sámi, the relatively high frequency of ambipositions in North Sámi, and the differential use of position for ambipositions in Finnish and Estonian, we advance three hypotheses that we will test in our study:

(1)   We expect to find regional variation in use of ambipositions since North Sámi is in contact with Norwegian/Swedish (predominantly prepositional) in Central and Western regions, and in contact with Finnish (predominantly postpositional) in the East.
(2)   We expect position to be associated with differences in expression of meaning.
(3)   We expect that a language with more ambipositions will use position in a more complex way; thus North Sámi should show more complexity than Finnish and Estonian, which should in turn be more complex than a language like Russian.

This study will focus on four North Sámi ambipositions: *miehtá* 'over', *čađa* 'through', *rastá* 'across', and *maŋŋel* 'after', here illustrated in use as both prepositions and postpositions (note that all adpositions govern the genitive case = GEN in North Sámi; these collocations are extracted from attestations in our database):

(4) a.   miehtá   dálvvi
          over   winter.GEN
              'during the winter'

    b.   dálvvi   miehtá
         winter.GEN   over
              'during the winter'

(5) a.   čađa   áiggi
         through   time.GEN
              'through time'

    b.   áiggi   čađa
         time.GEN   through
              'through time'

(6) a. rastá   joga
        across river.GEN
            'across the river'

    b. joga   rastá
        river.GEN   across
            'across the river'

(7) a. maŋŋel   soađi
        after   war.GEN
            'after the war'

    b. soađi   maŋŋel
        war.GEN   after
            'after the war'

A variety of criteria were considered in selecting these four ambipositions for the study. The initial group of candidates for the study were selected on the basis of frequency: we restricted the study to ambipositions that were of relatively high frequency so that it would be possible to apply statistical analysis to the behavior of ambipositions in pre- vs. postposition. To this end, we selected ambipositions that would yield at least 100 examples in our newspaper corpus (described below) for each position. All four ambipositions fulfilled or exceeded this frequency criterion. In addition we designed the selection to facilitate meaningful comparisons with Finnish and Estonian. This meant that it was best to select ambipositions that could express both time and space. The examples in 4–5 above illustrate the use of two of our ambipositions, *miehtá* 'over' and *čađa* 'through', to express time; in addition these ambipositions can express spatial relations, as illustrated in 8–9 (extracted from attestations in our database):

(8) a. miehtá máilmmi
        over   world.GEN
            '(all) over the world'

    b. turistabálgá   miehtá
        hiking-trail.GEN   over
            'along the hiking-trail'

(9) a. čađa   Ruota
        through   Sweden.GEN
            'through Sweden'

    b. vuovdde   čađa
        woods.GEN   through
            'through the woods'

The remaining two ambipositions are restricted to expression of only one domain: space only for *rastá* 'across', and time only for *maŋŋel* 'after' and could be thought of as 'control' ambipositions in contrast with the previous two that operate in both domains. Furthermore, the

ambipositions that express spatial relationships cover a range of different dimensions in their spatial expression: *rastá* 'across' characterizes a single dimension (typically crossing borders and rivers), *miehtá* 'over' characterizes two dimensions (typically surfaces and regions), while *čađa* 'through' characterizes three dimensions (typically volumes and conduits). Thus the four ambipositions in this study were judged likely to provide enough data for statistical analysis, to represent a variety of meanings, and to facilitate comparison with Finnish and Estonian.

The examples in 4–9 give the impression that position is entirely arbitrary, since both positions are attested, often with the same words. However, despite the presence of considerable flexibility, there are also strong asymmetrical tendencies. Our strategy is to establish the radial category networks of meanings that are relevant for each ambiposition and show how the distributions of attestations differ in relation to various factors such as prepositional vs. postpositional use, geography, and genre. In order to make this research possible, we assembled several corpora and databases, as detailed in the following section.

2. DATA AND METHODOLOGY. For the purpose of this study we put together various kinds of corpus material, including literary works, newspapers, and the New Testament. The literary works represent authors from three regions where North Sámi is spoken:

- Western region: Southern Troms in Norway and Jukkasjärvi in Sweden, where North Sámi is a minority language under strong pressure from Norwegian and Swedish
- Central region: Kautokeino in Norway, where North Sámi is not under as strong pressure from other languages
- Eastern region: Along the Tana River that forms the border between Norway and Finland, where there is pressure from Finnish

The translation of the New Testament is a recent work in which a deliberate effort was made to represent the language continuum of North Sámi and establish a normative standard (Magga 2004:52). Altogether 652 sentences containing the four ambipositions in our study were extracted from literary works and the New Testament, and all these examples were analyzed manually.

Texts from three newspapers, *Min Áigi*, *Áššu*, and *Ávvir*, representing publications for the years 1997–2011 were compiled into a corpus of 10 million words. The majority of newspaper journalism in North Sámi is undertaken in Norway, often using Norwegian texts as sources, so one would expect the use of language in newspapers to reflect Norwegian influence. Our corpus contained a total of 7,496 examples of our ambipositions. A minimum of 100 examples was analyzed by hand for each ambiposition in each position, yielding a total of 901 sentences. Table 1 shows the distribution of data analyzed in this study.

ambipositions	preposition		postposition	
	newspapers	literature + NT	newspapers	literature +NT
miehtá	133	72	100	25
čađa	102	34	158	99
rastá	101	37	100	56
maŋŋel	107	88	100	243

TABLE 1. Distribution of examples that were analyzed manually.

All of the example sentences and their analyses are publicly available at http://giellatekno.uit.no/adp/. This site also houses our statistical data and the code that was used to analyze this data by means of the R software package. In the remainder of this article we describe how this data was used to test the three hypotheses presented in the conclusion.

3. CONFIRMATION OF HYPOTHESIS 1: REGIONAL VARIATION. Hypothesis 1 states that we expect to find a stronger tendency to use ambipositions as prepositions in places where North Sámi is in contact with Norwegian and Swedish, but the opposite tendency in places where North Sámi is in contact with Finnish. Therefore we should expect to see more prepositional use in the Western (S. Troms) region, and more postpositional use in the Eastern (Tana) region, with the Central (Kautokeino) region falling somewhere between the two. Figure 1 represents the relative distribution according to region and genre, with all data from all four ambipositions aggregated. The black bars show the percentage of prepositional (PR) use, and the gray bars show the percentage of postpositional (PO) use. The first three clusters of bars compare the distributions of position in literary works across the three regions. These distributions align precisely with our expectations since prepositions predominate in the Western region, postpositions predominate in the Eastern region, and the Central region shows a more even balance, though with some preference for postposition. The regional differences are statistically significant with a robust effect size (chi-square = 129.7, df = 2, p<2.2e-16, Cramer's V = 0.48). Note that the distribution in the New Testament seems to fall between the Central and Eastern regions, whereas the newspapers show a distribution very similar to that in the West.
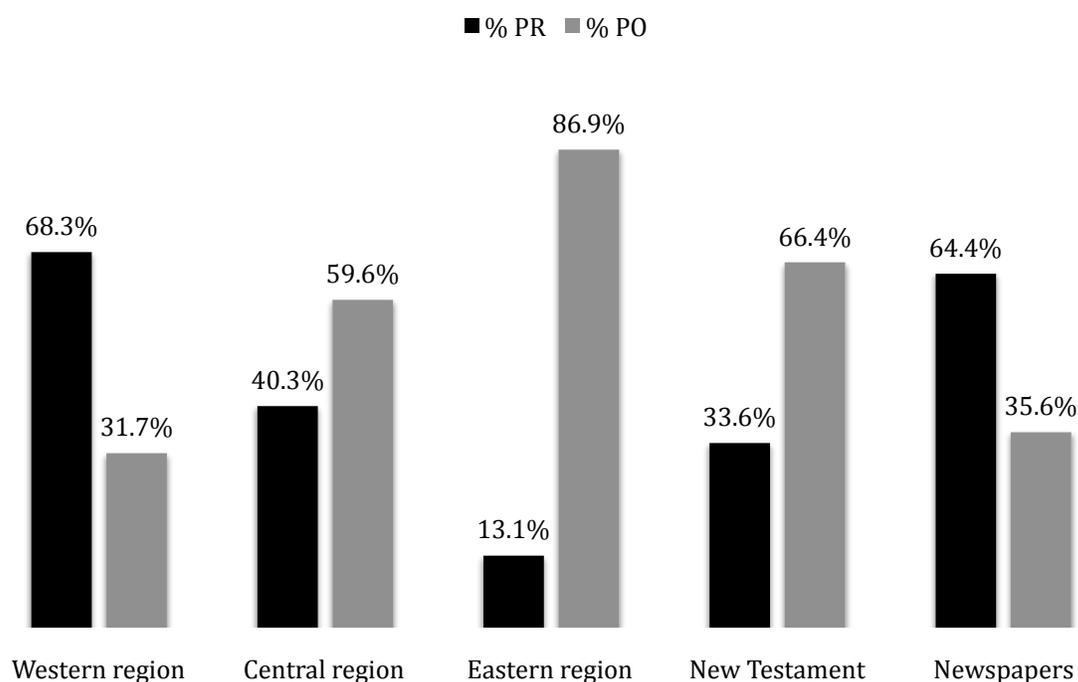
FIGURE 1. Distribution of prepositional vs. postpositional use.

Thus we can confirm hypothesis 1 concerning regional variation in use of position, and in addition we find differences in genre that likely stem from regional differences.

4. Confirmation of hypothesis 2: position and meaning. Hypothesis 2 states that we expect there to be a connection between position and the expression of meaning. We used the method of radial category profiling (Nesset et al. 2011, Janda et al. 2013) in order to test this hypothesis. Radial category profiling has been developed in order to measure differences between two (or more) highly synonymous polysemous linguistic units. The idea is that if two units share a single radial category, one can collect data on the distribution of attestations of the various meanings in the radial category and determine whether there are statistically significant differences. Often one finds that although the same range of meanings is attested for each of the synonyms, the center of gravity in the radial category is different for each one. For the purpose of this test we treated the prepositional vs. postpositional use of each ambiposition as a pair of near-synonyms.

The manual analysis of examples involved sorting them according to the meanings expressed. Based on the attested uses, a radial category network of meanings was established for each ambiposition. Three of the ambipositions, *miehtá* 'over', *čađa* 'through', and *rastá* 'across' were found to be polysemous, with 3–4 related meanings. It was discovered that in all three cases, the overall radial category of meanings expressed in prepositional and postpositional use was the same, but the distribution was quite different, with some meanings being more strongly associated with prepositional use and others more strongly associated with postpositional use. These differences were found to be statistically significant with robust effect sizes for the data collected from newspapers. Although the other data was too sparse for statistical analysis, it appears to differ somewhat showing an overall tendency to prefer expression of concrete spatial relations. The fourth ambiposition, *maŋŋel* 'after', was found to be monosemous and thus did not lend itself to this kind of analysis. We take up each of the first three ambipositions in turn in the following three subsections.

4.1. Radial category profiling of *miehtá* 'over'. Three meanings were identified in the radial category network of *miehtá* 'over': motion, extent, and time, as illustrated in 10–12 (the relevant adpositional phrase is boldfaced in both the original example and in the translation):

MOTION
(10)   Mii vánddardit **miehtá suohkana**, ja jearahallat olbmuid.
             'We wander **around the municipality** and interview people.'
EXTENT
(11)   Dat leat beaivelottit ja gávdnojit **miehtá máilmmi**.
             'Those are butterflies and they are found **all over the world**.'
TIME
(12)   Guovža oađđá **dálvvi miehtá**.
             'The bear sleeps **through the winter**.'

The three meanings of *miehtá* 'over' are closely related to each other. Both motion and extent describe a relation to a two-dimensional plane in the domain of space, but with motion we see a dynamic relation whereas the same relation is rendered as static with extent. In these two meanings the substantives that collocate with *miehtá* 'over' clearly refer to places that are conceived of as surfaces, such as the names of countries and other regions, the world, yards, and floors. The relation signaled by *miehtá* 'over' is rendered in the domain of time mostly with reference to the major portions of the diurnal and yearly cycle, which are apparently understood

as two-dimensional temporal objects: *dálvi* 'winter', *geassi* 'summer', *beaivi* 'day', and *idja* 'night'.

		MOTION	EXTENT	TIME
newspapers	Preposition (N = 133)	12%	79%	9%
	Postposition (N = 100)	0%	5%	95%
literature + NT	Preposition (N = 72)	10%	68%	22%
	Postposition (N = 25)	16%	36%	48%

TABLE 2. Radial category profiling of *miehtá* 'over'.

  Table 2 presents the radial category profiling analysis of our data on *miehtá* 'over'. We find strong differences in the distribution of the meanings of the ambiposition according to position. As a preposition it is primarily associated with EXTENT, whereas it mostly expresses TIME as a postposition. This difference is starker in the case of the newspaper data. There also appear to be differences between the genres, suggesting that MOTION is expressed more in literature and the New Testament, however the data from the latter is sparser.
  Differences in distribution can also be tracked at the level of the substantives that collocate with *miehtá* 'over'. For example, we have 27 attestations of prepositional use with *máilbmi* 'world', but no attestations of postpositional use with this noun. In the domain of TIME, in prepositional use there are only 8 attestations with *geassi* 'summer' and 5 with *dálvi* 'winter', but in postpositional use 26 with *geassi* 'summer' and 21 with *dálvi* 'winter'.

4.2. RADIAL CATEGORY PROFILING OF *ČAĐA* 'THROUGH'. Four meanings were identified in the radial category network of *čađa* 'through': MOTION, EXTENT, TIME, and MEANS, as illustrated in 13–16:

MOTION
(13) Gugán njuikii čuožžut ja viehkalii **vuovdde čađa** joksan dihte Katriinna.
    'Gugan jumped up and ran **through the forest** in order to catch up with Katrina.'
EXTENT
(14) ... oidnen ahte **čađa náhki** lea ráigi
    '... I saw that there was a hole **through the skin**'
TIME
(15) **Čađa áiggi** almmuhuvvojit ođđa girjjit dán davviriikkalaš dramatihka titána birra.
    '**Through time** new books were published about the drama of the titan of the North.'
MEANS
(16) **Skuvlla čađa** oažžu kultuvrralaš vuođu viidásit oahpuide.
    '**Through school** one gets the cultural basis for broader knowledge.'

The meanings of *čađa* 'through' are likewise related to each other and tend to reference spatial or temporal objects that are conceived of as three-dimensional. Forests, marketplaces, and towns, as well as weather conditions are common for both MOTION and EXTENT, and they are characterized by having both horizontal and vertical dimensions. Alternatively we find openings such as holes, doors, gates, and windows. Various kinds of problems and feelings are associated with metaphorical uses in these meanings. The temporal nouns associated with *čađa* 'through' are very different from those found with *miehtá* 'over'. Here we find two words that are very

frequent: *áigi* 'time' and *gaska* 'distance, interval' (always interpreted metaphorically to express a simultaneous action 'all during the time that ...'); very few other words are found here, though examples are *buolva* 'generation' (used in plural) and *jahki* 'year'. In addition to the domains of time and space, *čađa* 'through' can express relations in the domain of purpose, as we see in the MEANS meaning. Here the most typical substantives that collocate with *čađa* 'through' refer to organizations such as *Sámediggi* 'Sámi Parliament' and media like *TV* 'TV'.

		MOTION	EXTENT	TIME	MEANS
newspapers	Preposition (N = 101)	37%	8%	55%	0%
	Postposition (N = 158)	49%	2%	25%	25%
literature + NT	Preposition (N = 34)	65%	24%	12%	0%
	Postposition (N = 99)	94%	3%	2%	1%

TABLE 3. Radial category profiling of *čađa* 'through'.

Table 3 presents the radial category profiling analysis of our data on *čađa* 'through'. In the newspaper corpus TIME dominates the prepositional use of this ambiposition, while MOTION is the most frequent meaning expressed in postpositional use. In both genres, MEANS is found only in postpositional use. It appears that MOTION dominates use in both positions in literature and the New Testament, but our data on prepositional use is relatively sparse.

At the level of the substantive, we find some interesting trends. *Vuovdi* 'forest' and synonyms like *rohtu* 'thicket' signal difficult terrain, found only 6 times in prepositional use, but 21 times in postpositional use. Weather that is difficult to move through is found in only two attestations with prepositional use, one with *biegga* 'wind' (in plural), and one with *guoldu* 'snow flurry' (also in plural), but in postpositional use there are 14 attestations with these two words and near-synonyms. Metaphorical hindrances appear as *váivi* 'problem' (usually plural) and near-synonyms only 4 times in prepositional use, but 20 times in postpositional use. The different distribution of nouns referring to time is particularly interesting. As mentioned above, there are only two nouns that are relatively frequent here, namely *áigi* 'time' and *gaska* 'interval'. At first glance it seems that *áigi* 'time' is distributed equally across the two positions, since we have 27 attestations for each. However, the uses differ according to number: in prepositional use 24 attestations are singular as opposed to 3 that are plural, whereas the distribution is reversed in postpositional use where we find only 4 singular uses as opposed to 23 in the plural. The preference for the plural of *áigi* 'time' with postpositional *čađa* 'through' reflects the same pattern seen with forests and weather and problems. *Gaska* 'interval', however appears only in the singular and only in prepositional use, where it is attested 31 times. This one noun, *gaska* 'interval', is thus alone responsible for the most of the significant difference found in the expression of temporal vs. spatial meanings of *čađa* 'through' across prepositional and postpositional uses. However, the collocation *čađa gaskka* 'all the while' is attested only in our newspaper corpus. This particular collocation seems to be on its way to becoming fused into an adverb and can be often found written as one word, *čađagaskka*, on the internet.

4.3. RADIAL CATEGORY PROFILING OF *RASTÁ* 'ACROSS'. Three meanings were identified in the radial category network of *rastá* 'across': MOTION, EXTENT, and ENDPOINT, as illustrated in 17–19:

MOTION
(17)   Nu guhká go Guovdageainnu eatnu ii dulvva, de lea álki beassat **rastá eanu**.
              'As long as the Kautokeino river doesn't flood, it is easy to get **across the river**.'
EXTENT
(18)   Suohkan áigu boares telegráfastoalppuiguin ráhkadit šaldiid **muhtin jogaid ja jekkiid rastá**.
              'The municipality will use old telegraph poles to make bridges **across some rivers and marshes**.'
ENDPOINT
(19)   Rasmussen lohká sin áinnas viiddidit barggu **rájá rastá**.
              'Rasmussen says they would like to expand the work **on the other side of the border**.'

All of the meanings expressed by *rastá* 'across' reference the domain of space. Here we see both dynamic MOTION and static EXTENT, which we found also with *miehtá* 'over' and *čađa* 'through', plus a third spatial meaning: ENDPOINT. This third meaning is related to the other two by means of 'endpoint metonymy' (Janda 2010), where only the endpoint of a path is relevant. The relationship between MOTION and ENDPOINT is seen in English *over* in examples like *Sally walked over the hill* vs. *Jane lives over the hill*, where the latter involves only the endpoint. Similarly in example 19, no one is moving across the border, nor is the work stretched across the border, it is simply located on the other side of the border. In all three meanings *rastá* 'across' is often associated with objects that are conceived of as one-dimensional lines such as *johka* 'river', *rádji* 'border', *luodda* 'road', although wide bodies of water such as *mearra* 'sea' and *jávri* 'lake' are also found.

		MOTION	EXTENT	ENDPOINT
newspapers	Preposition (N = 101)	77%	23%	1%
	Postposition (N = 100)	48%	45%	7%
literature + NT	Preposition (N = 37)	92%	8%	0%
	Postposition (N = 56)	84%	14%	2%

TABLE 4. Radial category profiling of *rastá* 'across'.

MOTION predominates in both positions and in both genres, but is consistently stronger in prepositional use than in postpositional use, where in the newspaper corpus we find almost an equal portion of MOTION and EXTENT. MOTION appears to be relatively stronger in literature and the New Testament, but the data here is sparse.

Some nouns seem to show little or no preference for position with *rastá* 'across': for example, *johka* 'river' and its near-synonyms appear 36 times in prepositional use and 39 times in postpositional use, similarly *jávri* 'lake' and its near-synonyms appear 10 times in prepositional use and 12 times in postpositional use. However, *luodda* 'road' and its near-synonyms strongly prefer prepositional use with 29 attestations in that position as opposed to

only 7 for postpositional use. We see the opposite trend with *rádji* 'boundary', which appears 24 times in prepositional use but 47 times in postpositional use.

4.4. SUMMARY OF RADIAL CATEGORY PROFILING ANALYSIS. The radial category profiling analysis shows that all three of the ambipositions have different tendencies for both expression of meaning and collocation with specific nouns according to position. In other words, different meanings and nouns are characteristic for prepositional use than for postpositional use. The differences in tendencies are strongest for *miehtá* 'over' and *čađa* 'through' than for *rastá* 'across', but are significant for all three ambipositions, at least in the case of the newspaper data (see tests of significance and effect size on http://giellatekno.uit.no/adp/). Thus hypothesis 2 is confirmed: position in the use of North Sámi ambipositions is sensitive to the meaning expressed.

5. CONFIRMATION OF HYPOTHESIS 3: TYPOLOGY OF AMBIPOSITIONS. Hypothesis 3 expresses the expectation of a positive relationship between the extent of use of ambipositions in a language and the complexity of use of position with respect to meaning. Here we compare three types of languages:

- languages with minimal use of ambipositions — here our example is Russian
- languages with systematic use of ambipositions, representing 10–15% of adpositions – here our examples are Finnish and Estonian
- languages with systematic and extensive use of ambipositions, representing over 20% of adpositions – here our example is North Sámi.

Hypothesis 3 would lead us to expect the greatest complexity in the use of position with North Sámi, the least complexity with Russian, and Finnish and Estonian should fall somewhere between the two.

We extracted data on the three Russian ambipositions *spustja* 'after', *pogodja* 'after', and *radi* 'for the sake of' from the Russian National Corpus (http://ruscorpora.ru/) which contains over 200 million words. We found 395 attestations of *spustja* 'after', of which 243 represented prepositional use and 152 represented postpositional use. 924 attestations of *pogodja* 'after' were found; of these, about 10% were adverbial uses, 5% were prepositional uses, and the remainder were postpositional uses. *Radi* 'for the sake of' was much more frequent, with 18,137 attestations as a preposition and 7,304 as a postposition. Data was annotated for various possible factors, but we were unable to discover any differences connected to the expression of meaning in prepositional vs. postpositional use of the Russian ambipositions. The only trend we could find was that *radi* 'for the sake of' appears to prefer prepositional use when collocated with an animate noun, as in *radi detej* 'for the sake of the children', but postpositional use when collocated with an inanimate noun, as in *spravedlivosti radi* 'for the sake of fairness'.

For Finnish and Estonian, we rely upon secondary sources cited in section 1 (Huumo 2013, Lehismets 2011, Erelt 2003:117–118, Grünthal 2003). They report that position is indeed sensitive to meaning in the use of ambipositions in those languages, and furthermore that there are consistent tendencies across ambipositions, such that prepositional use tends to be associated with temporal expression whereas postpositional use is associated with spatial expression.

Our data shows that North Sámi also uses position in the expression of different meanings, but here we see more complexity since we do not find a consistent trend across

ambipositions. Both *miehtá* 'over' and *čađa* 'through' use position differently in relation to temporal vs. spatial expression, but the trends are opposed: *miehtá* 'over' prefers postpositional use in temporal expression, whereas *čađa* 'through' prefers prepositional use in temporal expression. Furthermore, we see a quite complicated picture at the level of the nouns that collocate with the ambipositions, with strong individual preferences.

The typological expectation in hypothesis 3 is confirmed. Russian, a language with few ambipositions, makes minimal or no distinctions with relation to position. Finnish and Estonian have systematic use of ambipositions and make consistent use of position to express meaning. The use of ambipositions in North Sámi is approximately double that in Finnish and Estonian and is also more complex, with different ambipositions showing different preferences for the use of position.

6. CONCLUSION. In this empirical study of data from newspapers and literary texts we show that North Sámi makes systematic and complex use of position to express meaning in collocations with ambipositions. It appears that the complexity of use of position is positively correlated to the extent that ambipositions are used in a language, with zero or little complexity in a language like Russian with few ambipositions, some complexity in a language like Finnish and Estonian with systematic use of ambipositions, and more complexity in a language like North Sámi with more extensive use of ambipositions. The use of position is also strongly influenced by language contact in North Sámi, where contact with Norwegian and Swedish is associated with a preference for prepositional use, but contact with Finnish is associated with postpositional use.

REFERENCES

BARTENS, RAIJA. 1974. Inarinlapin, merilapin ja luulajanlapin post- ja prepositiorakenteiden käytöstä. *Journal de la Société Finno-Ougrienne* 73.156–167.

ERELT, MATI (ed.). 2003. *Estonian Language.* Tallinn: Estonian Academy Publishers.

GRÜNTHAL, RIHO. 2003. *Finnic adpositions and cases in change.* Helsinki: Société Finno-Ougrienne.

HAGÈGE, CLAUDE. 2010. *Adpositions.* Oxford: Oxford University Press.

HUUMO, TUOMAS. 2013. Many ways of moving along a path: What distinguishes prepositional and postpositional uses of Finnish path adpositions? Turku: University of Turku, MS.

JANDA, LAURA A. 2010. Cognitive Linguistics in the Year 2010. *International Journal of Cognitive Linguistics* 1.1–30.

JANDA, LAURA A.; ANNA ENDRESEN; JULIA KUZNETZOVA; OLGA LYASHEVKAYA; ANASTASIA MAKAROVA; TORE NESSET; and SVETLANA SOKOLOVA. 2013. *Why Russian aspectual prefixes aren't empty: prefixes as verb classifiers.* Bloomington: Slavica Publishers.

KARLSSON, FRED. 2008. *Finnish. An Essential Grammar* (2nd ed.). London: Routledge.

LANGACKER, RONALD W. 2008. *Cognitive Grammar: A Basic Introduction.* Oxford: Oxford University Press.

LEHISMETS, KERSTEN. 2011. The role of verb semantics in the interpretation of Finnish path adposition expressions. Tartu: University of Tartu, MS.

MAGGA, TUOMAS. 2004. Dárogiela preposišuvdnaráhkadusaid jorgaleapmi. *Sámi dieđalaš áigečála* 1.43–53.

NESSET, TORE; ANNA ENDRESEN; and LAURA A. JANDA. 2011. Two ways to get out: Radial Category Profiling and the Russian Prefixes vy- and iz-. *Zeitschrift für Slawistik* 56.377–402.

NICKEL, KLAUS PETER, and PEKKA SAMMALLAHTI. 2011. *Nordsamisk grammatikk.* Karasjok: Davvi Girji AS.

NIELSEN, KONRAD. 1979 (1926–1929). *Lærebok i lappisk (samisk). Bind I: grammatikk.* Oslo: Universitetsforlaget.

TAYLOR, JOHN R. 2003. *Linguistic Categorization.* Oxford: Oxford University Press.

YLIKOSKI, JUSSI. 2006. Fuomášumit sámegiela adposišuvnnaid funkšuvnnain: ovdamearkan *alde-* ja *ala-*postposišuvnnaid ii-lokála geavaheapmi. *Sámi dieđalaš áigečála* 1.39–61.

YLIKOSKI, JUSSI. 2009. *Non-finites in North Saami.* Helsinki: Société Finno-Ougrienne.

The emergence of syllable structure?
Data from gradient vowel reduction in Brazilian Portuguese

Ricardo Napoleão de Souza
Universidade Federal de Minas Gerais / University of New Mexico

ABSTRACT. Vowel reduction may be the first step towards the emergence of new syllable patterns in a language. Reduction is understood here as the time and gestural compression of phonetic units, which may or may not lead to vowel deletion. Reduction in words containing high vowels [i] and [u] in unstressed CVC-type syllables was evaluated. The data show that ultimately deletion occurs at high rates, yielding novel consonant clusters in the language. Usage-based phonology (Bybee 2001, 2010) provides the theoretical framework to explain how gradual phonetic change may lead to phonological change.

1. INTRODUCTION. Vowel reduction may be the first step towards the emergence of new syllable patterns in a language. Reduction is understood here as the time and gestural compression of phonetic units which may or may not lead to vowel deletion. The assumption that phonological change is a gradient processes is supported by studies based on dynamic phonological theories such as USAGE-BASED PHONOLOGY (Bybee 2001, 2010) and ARTICULATORY PHONOLOGY (Browman & Goldstein 1986, 1992). Dynamic phonological theories predict that phonological change derives from gradual changes on lexical and phonetic levels, thus affecting different words and phonetic environments in distinct ways.

Gradience in vowel reduction phenomena provides evidence that phonetic detail is actually present in phonological representations, suggesting a greater proximity between levels traditionally regarded as distinct. For instance, speakers can process (and produce) the same word with different vowel qualities, degrees of voicing and vowel durations (including zero duration, i.e. vowel deletion) every time they speak. Each of these different pronunciations is stored in the speaker's memory, thus enabling processing of gradually variable vowel durations. Hence, phonetically different instances of the same unit can be naturally processed as part of language representation without eliminating redundant information.

In this context, articulatory phonology (Browman & Goldstein 1986, 1992) is a useful framework to explain variation data, for it accounts for phonetic gradience as a result of articulatory adjustments. In articulatory phonology, gestures are the basic units of phonological contrasts. Gestures are abstract characterizations of articulatory events that have intrinsic durations. According to Iskarous (2010), important principles of articulatory phonology are:

- Dynamic principles and adjustable motor parameters explain sound variation.
- Various aspects of phonological structure result from the dynamic expression of linguistic tasks and their interaction.

A great deal of phonetic and phonological variation are found in Brazilian Portuguese. In the Belo Horizonte variety, an ongoing vowel reduction phenomenon can be observed in the pronunciation of pre-stressed CVC syllables in words containing high vowels. Two examples of the variation are given below.

1. Reduction of [i]: pistola [pis]tola ~ [pi̥s]tola ~ [ps]tola: 'pistol';
2. Reduction of [u]: buscava [bus]cava ~ [bu̥s]cava ~ [bs]cava: '(s/he)sought'

This study investigated the reduction of pre-stressed high vowels [i] and [u] in the Belo Horizonte dialect of Brazilian Portuguese. The two vowels were examined in the context of CVC-type syllables in which the coda consonant position was always occupied by the sibilant [s].

2. VOWEL ARTICULATION AND REDUCTION

2.1. ARTICULATORY CHARACTERIZATION: PHONATION. An important factor for evaluating the influence of voicing gradience in vowel reductions is to understand the different configurations of the vocal cords during speech production. Vowel segments are articulated by means of the vibration of vocal cords, which are stretched or relaxed due to larynx muscle activity (Martin 2008). The larynx either allows the sound to flow freely through glottal space to produce voiceless sounds or it can adjust the vocal cords to disturb the airflow to produce different degrees of voicing. Voicing can be understood as the product of the rapid opening and closing movements of the glottis, in response to air pressure generated in the trachea. The muscular action of the larynx can be controlled by the speaker, who may adjust the vocal cords in order to produce phonatory effects beyond voicing and devoicing.

Just as the muscular action of the articulators is variable and presents different degrees, it is expected that voicing too can be gradual. Variations in control of the glottis can be employed in different languages to indicate linguistic information, especially through phonation contrasts between two or more types of vowel (Ladefoged 1971).

A criterion commonly used to characterize vowel production phonetically is to specify the different configurations of glottal space during speech. Ladefoged (1971) suggests that different kinds of phonation be defined by the degree of openness of the vocal cords. Phonatory differences can therefore be classified on a voicing continuum, ranging from total glottal occlusion through devoicing to modal voicing and beyond (Gordon & Ladefoged 2001). The diagram below represents phonation types according to the degree of openness of the vocal cords.

Most open ←				→ **Most closed**
Phonation type Voiceless	Breathy	Modal	Creaky	Glottal closure
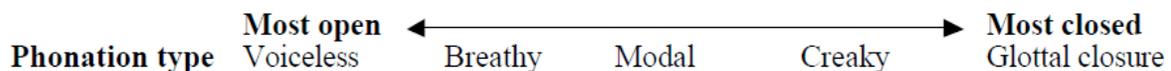
FIGURE 1. Phonation Types According to Glottal Configuration
(Gordon & Ladefoged 2001).

The diagram shows that the gradual nature of phonatory types allows for vowel articulation to be expressed in a gradient manner, taking the form of a voicing continuum. The existence of a voicing continuum thus implies phonetic gradience in the production of vowels. An articulatory examination of vowel production in contexts of language change and variation may indicate a gradual alteration of voicing in vowel reduction phenomena. The voicing continuum serves as an articulatory explanation for phonetically motivated vowel deletion in cases such as [pis]tola ~ [pi̥s]tola ~ [ps]tola for *pistola* 'pistol'.

Articulatory production of vowels was examined in this research by means of an electroglottograph. This electric device measures the vibrations of the vocal cords via electrodes placed around the speaker's larynx area (Rothenberg & Mashie 1988). The electroglottograph (EGG) produces a physiologically safe electric current that monitors vocal cord movement during speech, allowing for the characterization of vowels in an articulatory gradient manner (Vieira 1997).
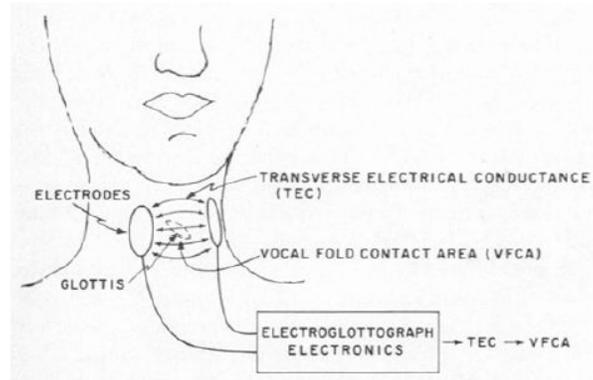
FIGURE 2. Schematic representation of how the EGG words (Rothenberg & Mashie 1988).

During phonation, the vocal folds are intermittently separated by the glottis. As the vocal folds move apart, the glottis opens, increasing the electrical impedance across the larynx. When the vocal folds come closer together, the size of the glottis subsequently decreases the electrical impedance across the larynx. These changes in the electrical current are then recorded into the EGG to be later aligned with acoustic measures, such as Praat (Boersma & Weenik 2012) for determining voicing patterns. The figures below show how acoustic and articulatory data can be paired to allow for an analysis of gradualness in the voicing of vowels.

[p i s]
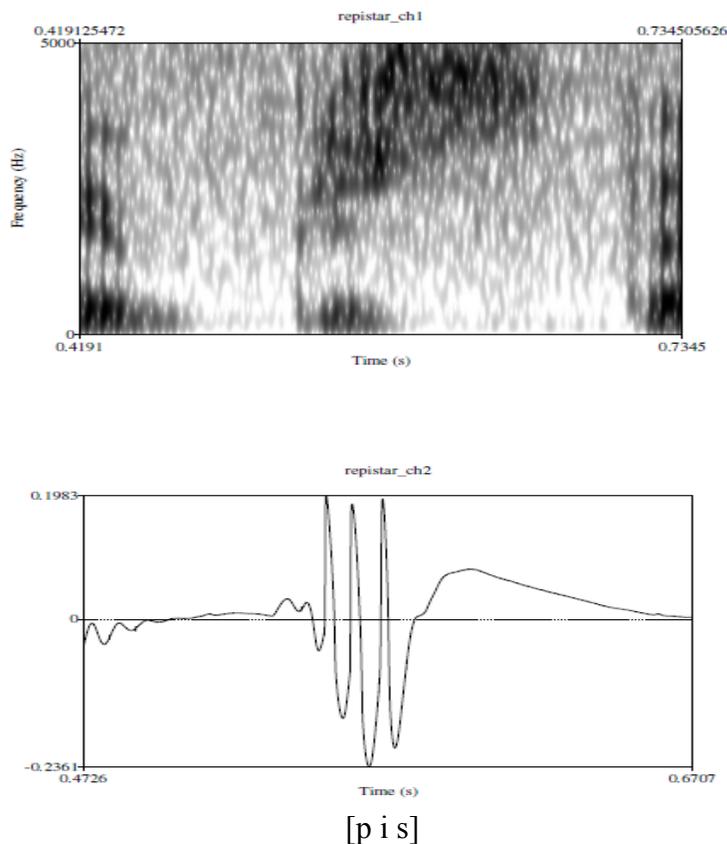
FIGURE 3. Spectrogram and EGG data paired for the syllable [pis].

2.2. ACOUSTIC IDENTITY: THE BRAZILIAN PORTUGUESE VOWEL INVENTORY. Word stress is variable in Portuguese. Stressed and unstressed vowels have different statuses in Brazilian Portuguese (henceforth BP). In BP, the number and distribution of vowels are directly related to stress placement. This pattern that differentiates vowels according to stress is linked to a number of historical vowel change phenomena that have given rise to the current vowel inventory. The present description is based on the early work of Câmara Jr. 1973 and a more recent characterization in Cristófaro-Silva 1996. Historical data come from Teyssier 2001.

In contemporary BP, seven oral vowels [a, e, ɛ, i, o, ɔ, u] contrast in stressed position, whereas in unstressed positions the number of contrasting vowels is smaller. In pre-stressed and post-stressed medial positions, five vowels [i, e, a, o, u] have vowel contrastive status. Word-finally, only three vowels [ɪ, ə, ʊ] contrast in most varieties of BP. These final unstressed vowels have arisen from reductions in unstressed word final position in more recent times: [i] > [ɪ], [a] > [ə] and [u] > [ʊ] (Cristófaro-Silva 1996). This is the case of the spoken language in Belo Horizonte.

Historically, close mid vowels have been affected by raising phenomena (Teyssier 2001), i.e. [e] > [i], [o] > [u] in unstressed final position in the majority of dialects. In addition, the contrast between the pairs of mid vowels [e, ɛ] and [o, ɔ] has long been lost in unstressed positions. The complete inventory of oral vowels in the Belo Horizonte variety of BP can thus be summarized as follows:

PRE-STRESSED AND POST-STRESSED VOWELS	STRESSED VOWELS	FINAL UNSTRESSED VOWELS
[i]	[i]	
[e]	[e]	[ɪ]
	[ɛ]	
[a]	[a]	[ə]
[o]	[o]	
	[ɔ]	[ʊ]
[u]	[u]	

TABLE 1. BP vowels according to stress and position within the word.

2.3. VOWEL DURATION IN BP. The present work uses vowel duration measurements as evidence of pre-stressed [i] and [u] reduction. Both internal and external factors influence the duration of a vowel. Important external factors are phonetic quality of the adjacent segments and position of the vowel within the word. This research tested the influence of both factors in duration. The results of three previous studies that measured the duration of vowel segments in BP are presented in Moraes 1999, Faveri 2000, and Cantoni 2013.

Although the data in Moraes 1999 and Faveri 2000 do not come from the variety under investigation in this paper, they serve as indications of how vowel duration influences the possible emergence of consonant clusters. I will show that shorter vowels in the Belo Horizonte dialect may suggest why this variety shows signs of new syllable patterns. Moreover, these three studies confirm the crosslinguistic tendency for high vowels to be shorter than mid or low vowels (Lehiste 1970, inter alia).

Moraes (1999) measured the duration of vowels in pre-stressed, stressed and unstressed final position in an experiment with speakers of the Rio de Janeiro variety of BP in three speech modes. The table below illustrates the durations of the vowels pronounced in regular speaking mode in milliseconds:

CONTEXT / VOWEL	i	e	ɛ	a	ɔ	o	u
PRE-STRESSED	80.3	91.6	-	97.6	-	94.6	79.6
STRESSED	125.6	136.4	150.5	141.6	150.7	140.2	127.6
POST-STRESSED	69.7	-	-	101.9	-	-	71.9

TABLE 2. Mean Duration of BP Vowels in Moraes 1999.

Table 2 shows that high vowels had shorter durations, and the vowel [i] is the shortest in pre-stressed and tonic contexts. In the pre-stressed context, there was a slight difference between [u] and [i], [u] being the shortest.

Faveri (2000) studied oral vowels in the Portuguese spoken in Santa Catarina state. The results point to a shorter duration of the vowel [i] among all the vowel inventory of the PB. As in Moraes 1999, the high vowels [i] and [u] are the shortest. The average durations obtained from her study were as follows:

CONTEXT / VOWEL	i	e	ɛ	a	ɔ	o	u
PRE-STRESSED	46,3	70,9	-	72,4	-	70,3	61,9
STRESSED	83,9	116,3	119,2	107,6	126,2	104,6	103,4
POST-STRESSED	49,7	-	-	64,0	-	-	54,8

TABLE 3. Mean Duration of BP Vowels in Faveri 2000.

Cantoni (2013) measured the duration of the vowels [i] and [u] as part of her dissertation on prosodic patterns of Belo Horizonte Portuguese. She gathered vowel duration data related to pre-stressed, stressed and final unstressed position. The following results were obtained:

CONTEXT / VOWEL	i	u
PRESTRESSED	39,7	51,9
STRESSED	75,9	86,4
POST-STRESSED	23,5	38,8

TABLE 4. Mean duration of [i] and [u] in Cantoni 2013.

Consistent with the results from Rio de Janeiro (Moraes 1999) and Santa Catarina (Faveri 2000) variants of BP, Cantoni's (2013) measurements of the durations of pre-stressed vowels [i] and [u] in Belo Horizonte show that they are shorter than those of the same vowels in stressed position. The data presented above reinforce the evidence for the intrinsic shorter duration of high vowels in all positions within the word in BP. The results in Cantoni 2013 also point to the fact that high vowels are comparatively much shorter in the Belo Horizonte variety than in other dialects of BP.

2.4. VOWEL DELETION AND THE EMERGENCE OF SYLLABLE PATTERNS. The non-production of vowels in syllables of the CVC-type would result in the emergence of unusual phonotactic patterns in the Portuguese of Belo Horizonte. The absence of a vowel in the syllable nucleus would generate a novel sequential arrangement of the remaining consonants, such as in *piscar* [pis]car > [psk]ar 'to blink'. In pronunciations like [psk]ar, the syllabic pattern expected for Portuguese, which consists of a maximum of two consonants in tautosyllabic position, would be violated in a sequence of three consecutive consonants. The maximum number of segments that make up a Portuguese syllable has the following configuration (adapted from Cristófaro-Silva 1996):

SYLLABLE		SEGMENT
	C_1	stop
	C_2	liquid
$C_1C_2VC_3$	V	vowel or diphthong
	C_3	sibilant or rhotic

TABLE 5. Longest syllable type in Portuguese.

Thus, the emergence of consonant clusters with three or possibly more consonants would be highly innovative, since Portuguese typically allows only sequences of stop consonants

followed by either one of two liquid consonants ([l, ɾ]) (Câmara Jr 1973, Parkinson 1990, Cristófaro-Silva 1996, Collischonn 2001). If high vowels are indeed deleted in CVC-type syllables, some of the innovative segment sequences would look like the following:

VOWEL	STANDARD PRONUNCIATION	NOVEL PRONUNCIATION	EMERGING SYLLABLE	GLOSS
[u]	[kus].tó.dia	[kstɔ].dia	*CCCV	'custody'
	[bus].car	[bskah]	*CCCVC	'to search'
[i]	[pis].ta.che	[psta].che	*CCCV	'pistacchio'
	[pis].car	[pskah]	*CCCVC	'to blink'

TABLE 6. Emerging syllables from vowel deletion.

High vowel deletion therefore has important implications both for syllable and phonotactic patterns in BP. Given that it is solely the vowel that is deleted in such instances, the remaining consonants would have to be regrouped in unusual combinations. The next section reviews a number of studies that correlate phonetic factors to vowel reductions.

3. VOWEL REDUCTIONS AND DELETION IN THE WORLD'S LANGUAGES. Gordon (1998) presents a survey of 55 languages in which non-modal vowels occur. Non-modal vowels are defined as any vowel sound produced with phonation types other than full, i.e. creaky voiced or devoiced vowels. Based on the analysis of phonetic and phonological phenomena affecting these languages, he cites a number of common features that lead to devoiced or voiceless vowels. Factors that trigger devoicing are commonly of articulatory (vowel height, duration, adjacent segments) and prosodic character (distance to prosodic border, stress placement) affecting mostly unstressed vowels.

In a personal communication on the typology of voiceless vowels, Chitoran and Marsico (2010) present a survey of the factors that favor phonetic vowel devoicing in 39 languages, among them BP. These factors coincide with those cited by Gordon (1998) and suggest a structural explanation for the occurrence of vowel devoicing. Both works suggest that there are restrictions to devoicing depending on the type of segments in the vowel's phonetic vicinity. The table below illustrates the five factors that motivate devoicing, according to Gordon (1998) and Chitoran & Marsico (2010):

CONDITIONING FACTORS

Vowel Height
Vowel Stress
Speech Rate
Voicing of Adjacent Segments
Distance to Prosodic Borders

TABLE 7. Conditioning Factors for Vowel Devoicing in Gordon 1998 and Chitoran & Marsico 2010.

If voicing is considered to be a continuum, devoicing is but a point on the scale that goes from the full articulation of the vowel to its total deletion in the opposite end. Such a voicing scale would therefore include deletion and devoicing in the same analysis. Consequently, deletion would be the final stage of a vowel reduction phenomenon in the direction of consolidating the change.

Previous studies show that high vowel devoicing takes place in BP. Câmara Jr (1973), Bisol and Hora (1993), Abaurre and Sândalo (2007), Leite (2007), Battisti and Herman (2008) and Assis (2010) all report devoicing of [i] in unstressed position. I showed in Napoleão 2010 that in addition to [i], there is a combination of structural factors that favor the devoicing of the vowel [u] in pre-stressed CVC syllables closed by the fricative [s]. In the next section the method used to investigate the correlation of vowel reduction, devoicing and deletion in the present research.

4. METHODOLOGY

4.1. CRITERIA FOR THE CLASSIFICATION OF OVERALL VOWEL REDUCTION. To test the hypothesis that high vowel reduction leads to vowel deletion, a controlled experiment was devised. This work specifically aimed to evaluate the behavior of high vowels in pre-stressed CVC syllables in words such as *pistache* [pis]tache 'pistachio' and *buscado* [bus]cado '(it was) sought' in Belo Horizonte Portuguese. The target vowels were classified as reduced and assigned the value 1 if the presented atypical spectrographic characteristics. Non-reduced (or full) vowels were attributed the value 0.

Cases in which there was no visible formant information, (i.e. cases of vowel deletion) were also labeled variant 1. The EGG was used to determine whether a vowel had been deleted or simply reduced. All vowels whose formants presented the distinct acoustic characteristics of the vowel in question were considered to be variant 0 - full vowel. The values were subsequently tallied for each individual token so as to assess the rate of reduction in the sample. Figure 4 below illustrates instances of vowels that were considered full (left) and reduced (right) in the sample, and Figure 5 illustrates a case of vowel deletion.
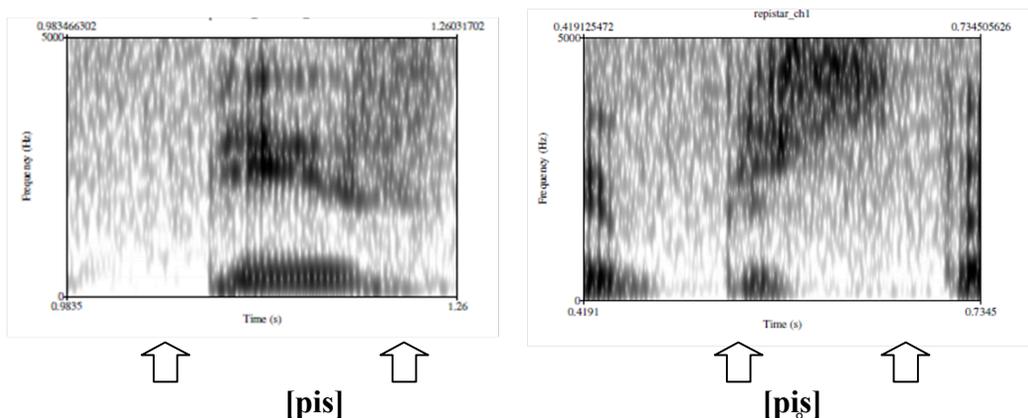
FIGURE 4. Spectrograms showing the syllable [pis] in full and reduced pronunciations in the words *copista* 'copier' and *episcopal* 'bishop's', respectively. (Speakers G3 and H2, slow repetition). The arrows indicate the approximate beginning and end of the syllables.

[pis]   **[ps]**
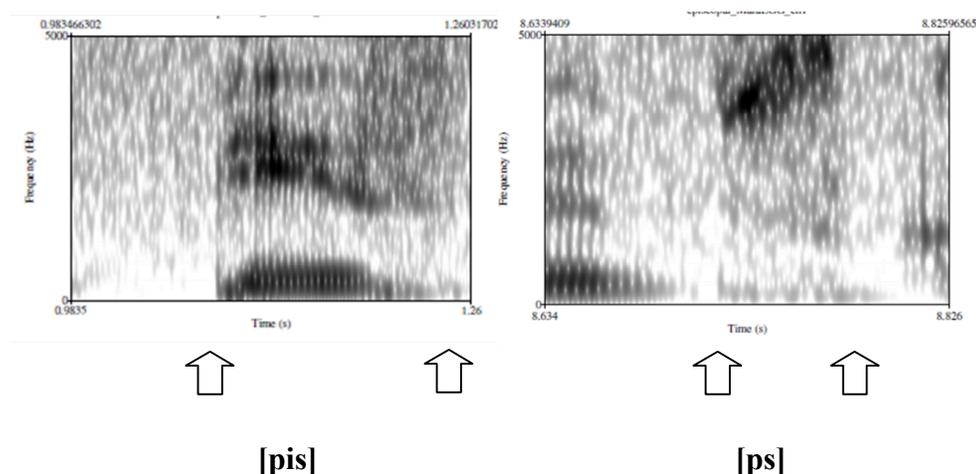
FIGURE 5. Spectrograms showing the syllable [pis] in full and reduced (deleted) pronunciations in the words *copista* 'copier' and *episcopal* 'bishop's', respectively. (Speakers G3 and G4, slow repetition). The arrows indicate the approximate beginning and end of the syllables.

Despite different methodological approaches, previous studies (cf. Section 3 above) on unstressed vowel reduction reviewed above show broad agreement on the structural factors that motivate the occurrence of the phenomenon. These observations were taken into account in the experiment design, which sought to verify their accuracy regarding Belo Horizonte Portuguese. The lexical items in the sample were thus assessed according to the following structural features:

- VOWEL FRONTNESS – [i] or [u]. Previous works indicate that the vowel [i] is more prone to undergo reduction phenomena than the back vowel [u].
- SPEECH RATE – faster or slower speech rate. Faster speech reportedly favors the occurrence of reduction, for it increases the chance for gestural overlap.
- VOICING OF THE PRE-VOCALIC CONSONANT IN THE SYLLABLE TESTED – voiced or voiceless prevocalic consonant. Voicing overlap can cause vowels occurring after voiceless consonants to undergo reduction by assimilation.
- POSITION OF THE SYLLABLE TESTED – initial or non-initial position within a word. Initial syllables show prominence that could hinder vowel reduction phenomena in their nuclei.
- TOKEN FREQUENCY – higher token frequency versus lower token frequency. Token frequency reportedly favors the occurrence of phonetically-conditioned reductions.

4.2. ORGANIZATION OF THE WORDS TESTED. In total, sixty words containing the sequence C₁V[s] were selected for the experiment. High vowels [i] and [u] were the dependent variables in this research. Lexical items were selected from two sources: the corpus of the Avaliação Sonora do Português Atual (ASPA) Project and the online Houaiss Portuguese Dictionary (Houaiss, Villar & Mello 2009). The main criterion for the selection of the items was the phonetic context, irrespective of the grammatical category to which the words belonged. Nouns were nonetheless preferred whenever possible. The CVC syllables found in the dictionary are shown below.

VOWELS	CONSONANTS					
	p_s	b_s	t_s	d_s	k_s	g_s
i	✓	✓	-	✓	-	-
u	✓	✓	✓	-	✓	✓

TABLE 5. CVC syllables containing [i] or [u] found.

Two experimental control measures were taken to control for stress and vowel height. To test if primary stress restricted the occurrence of reductions, words containing stressed CVC sequences containing [i] and [u] were added, for example *sambista* [sã.'bis.tə] 'samba dancer' and *robusta* [ho.'bus.tə] 'robust'. As a control factor for vowel height, words containing pre-stressed CVC sequences in which V was the low vowel [a], e.g. *pastagem* [pas.'ta.ʒẽɪ] 'pasture' were also included (see Table 6 below).

Phonological (prosodic) words consisting of a noun plus a different word (e.g. *o bistrô* 'the bistro') were used in cases where a word containing the CVC syllable configuration desired was not found. In addition, some words that have mid vowels subject to raising were used. Tables 6–7 below display the words used in the experiment. Control words are shown in italics, words containing vowels subject to raising are shown in brown shading and phonological words are shown in blue.

VOWEL / CONSONANT		[i] [p]	[i] [b]	[u] [p]	[u] [b]	[a] [p]	[a] [b]
STRESSED	+ FREQ.	cam.pis.ta	sam.bis.ta	ca.puz	ro.bus.ta	*pas.ta*	*bas.ta*
	- FREQ.	co.pis.ta	clu.bis.ta	pús.tula	bus.ta	*em.pas.te*	*bas.tos*
PRE-STRESSED	INITIAL + FREQ.	pis.tola / pis.tache	bis.coito / bis.caia	pos.tiço / pus pressão	bus.cando / bus.cava	*pas.tagem*	*bas.quete*
	INITIAL - FREQ.	pis.coso / pis.cada	bis.cate / bis.pado	pus.tulenta / pus.tuloso	bus.pirona / bus.tuário	*pas.toso*	*bas.tarda*
	MEDIAL + FREQ.	e.pis.copal / des.pis.tar	o bis.trô / o bis.turi	com.pos.tura / ca.puz preto	com.bus.tão / ro.bus.tez	*pas.coal*	*a.bas.tecer*
	MEDIAL - FREQ.	des.pis.tou / re.pis.car	ra.bis.tel / ra.bis.car	o.pus.cular / cre.pus.cular	re.bus.cada / ro.bus.tece	*re.pas.tar*	*a.bas.tado*

TABLE 6. Words with CVC syllables starting with [p, b] tested.

FREQUENCY	SYLLABLE TYPE			
	[dʒis]	[tus]	[kus]	[gus]
+ freq.	disputa	tostão	custeio	gustavo
- freq.	discado	tuscana	custódia	gustaio

TABLE 7. Words with CVC syllables starting with [dʒ, t, k, g] tested.

The token frequency of each item was checked using the ASPA Project frequency counter. The sample was then divided into two groups according to their individual frequency: more frequent items (total token frequency of over 70 in the corpus) and less frequent items (0 to 50 occurrences in the corpus). Frequencies were obtained from a corpus of 228,766,402 written words compiled by the ASPA project. A corpus of written language was selected due to the unavailability of a spoken language corpus of Belo Horizonte Portuguese at the time this research was undertaken.

4.3. EXPERIMENT DESIGN AND DATA COLLECTION. The present study was based on the analysis of speech data recorded in an experiment with 16 adult informants from Belo Horizonte, Brazil. To participate in the study, informants had to originally be from Belo Horizonte and to have always lived in the city with no absence longer than six months. All informants had a similar level of education; they were namely undergraduate students at the Federal University of Minas Gerais (UFMG) or already possessed a bachelor's degree. Two age groups were established: half of the participants were younger than 25 years old and half was older than 35 years old. The mean age of the younger group was 23 years and 48 years for the older group. An equal number of men and women participated.

The 60 words tested in the present study were put in different carrier sentences created for the purpose of assessing pre-stressed high vowel reduction. Carrier sentences contained a total of eight (8) syllables each, with adjacent syllables alternating between strong and weak to ensure prosodic systematicity. The syllable tested was always the third one in the sentence. These measures were introduced so as to control for prosodic prominence in the syllables tested. Carrier sentences varied according to the length of the word tested (two, three or four syllable-words). The overall number of syllables in the carrier sentences was kept the same. Although distance to prosodic borders does play a role in unstressed vowel reduction (Gordon 1998, Chitoran & Marsico 2010) the present study did not test such variables.

To assess whether speech rate influences reduction, subjects were asked to read each set of four repetitions with increasing speed. The first reading was to be at the speaker's normal reading speed. From the second reading on, informants were asked to read each sentence a little faster. The fourth and last reading was the fastest. The sentences were presented in slides programmed to change at predetermined time lengths. Visual aids helped the participants keep track of the speed of the repetition. Only the first and last repetitions, that is, the slowest and the fastest ones, were included in the present analysis.

Participants were recorded individually in a soundproof booth inside the Center of Speech Studies (CEFALA) of the Faculty of Engineering at the University of Minas Gerais (UFMG), in Belo Horizonte, Brazil. Recordings began after a short habituation phase in which the outline of the experiment was explained to the speakers. They were not informed of the goal of the experiment and took four programmed breaks during each session. Each participant could choose when s/he wanted to start the experiment again after the breaks.

5. RESULTS AND DISCUSSION. The recordings produced 1,920 tokens of vowels for the analysis. Of these, 108 were discarded due to several reasons, totaling in 1,812 vowel tokens analyzed in this study. The results of the overall vowel reduction analysis follow.

VOWEL	TOKENS	REDUCED TOKENS	%
[i]	562	346	62
[u]	584	205	35
TOTAL	1.176	551	47

TABLE 8. Overall results.

When broken down according to speech rate (fast versus slow) results showed a clear tendency for reduction in the fast speech rate, as predicted by articulatory phonology and usage-based theory. Results are shown in Table 9, and Figure 6 below shows the results in a graph. A chi-square test was used to evaluate the difference between the speech rates.

	[i]		[u]		BOTH		
	N	%	N	%	N	%	p
SLOWER	118	41	37	12	155	26	<0.0001
FASTER	228	79	168	55	396	67	

TABLE 9. Results according to vowel type and speech rate.
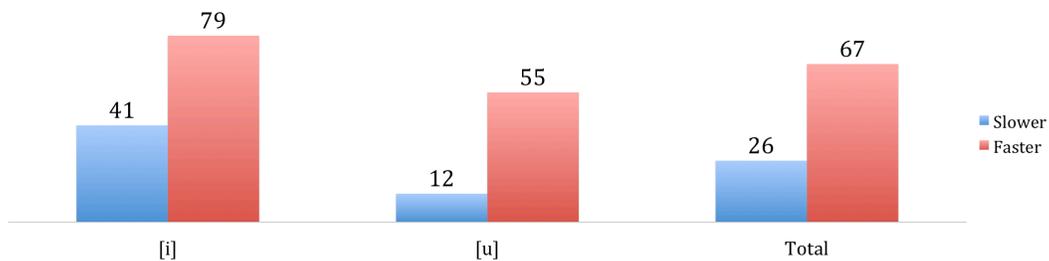
FIGURE 6. Results according to vowel type and speech rate.

As explained in the introduction, the term reduction is used in this paper to include both partially reduced (i.e. shorter) vowels and deleted vowels. If one breaks down the above results into those two categories, partially reduced versus deleted vowels, the following percentages emerge:

VOWEL	PARTIALLY REDUCED	DELETED
[i]	0.33	0.67
[u]	0.42	0.58

TABLE 10. Percentage of partial reductions versus deletion of [i] and [u].

The numbers show that deletions comprise most of the reduction cases for both vowels, suggesting not only that new consonant sequences are emerging but also that vowel deletion is not a categorical phenomenon. These facts corroborate predictions made by articulatory phonology and usage-based theories in that gradualness is an important part of phonological change.

6. CONCLUSION. Statistical analyses of the results demonstrate that reduction of pre-stressed [i] and [u] is taking place at high rates in the dialect. Moreover, the data showed that there is indeed a trend for vowel deletion, thus yielding unusual sequences of segments of up to three consonants. Words were affected differently according to a number of factors tested. Variables such as vowel type, speech rate, stress and speaker age were relevant in the occurrence of high vowel reduction. Frequency effects played a role in words containing [u]. The high rate of reductions for the vowel [i] might suggest it is no longer subject to frequency effects, given that the change is approaching completion. Additional experiments are needed to evaluate this hypothesis.

The experiment indicates that vowel reduction may lead to the emergence of new syllable patterns. Differences in the range of occurrence of the phenomenon suggest that this ongoing change manifests itself in a gradient fashion. Compensatory lengthening as well as the simultaneous occurrence of deletion and partial reductions contribute to the view that gradience plays a role in the emergence of new linguistic units.

REFERENCES

ABAURRE, MARIA, and FILOMENA SÂNDALO. 2007. Vowel deletion and reduction: acoustic clues of rhythmic organization in Brazilian and European Portuguese. Paper presented at PaPI 07 Phonetics and Phonology in Iberia, Braga, Portugal.

ASSIS, ALINE. 2010. *O apagamento de vogais pretônicas no POBH – norma culta*. Belo Horizonte, MG: Federal University of Minas Gerais (UFMG) thesis.

AVALIAÇÃO SONORA DO PORTUGUÊS ATUAL (ASPA) PROJECT. Online: http://www.projetoaspa.org/.

BATTISTI, ELISA, and BEN HERMANS. 2008. A palatalização das oclusivas alveolares: propriedades fixas e variáveis. *Alfa* 52. 279–288.

BISOL, LEDA, and DERMEVAL DA HORA. 1993. Palatalização da oclusiva dental e fonologia lexical. *Letras* 5.25–40.

BOERSMA, PAUL, and DAVID WEENIK. 2012. Praat: Doing phonetics by computer [computer program], version 42.07. Online: http://www.praat.org/.

BROWMAN, CATHERINE, and LOUIS GOLDSTEIN. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3.219–252. Cambridge: Cambridge University Press.

BROWMAN, CATHERINE, and LOUIS GOLDSTEIN. 1992. Articulatory phonology: an overview. *Phonetica* 49.155–180.

BYBEE, JOAN. 2001. *Phonology and language use.* Cambridge: Cambridge University Press.

BYBEE, JOAN. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

CÂMARA JR, JOAQUIM M. 1973. *História e estrutura da língua portuguesa*. Petrópolis: Editora Vozes.

CANTONI, MARIA. 2013. *O acento no português brasileiro*: *uma abordagem experimental*. Belo Horizonte, MG: Federal University of Minas Gerais (UFMG) dissertation.

CHITORAN, IOANA, and EGIDIO MARSICO. 2010. Vowel devoicing: An updated phonetic typology. *Berkeley Linguistics Society* 39.16–16.

COLLISCHONN, GISELA. 2001. A sílaba em português. *Introdução a estudos de fonologia do português brasileiro*, 91–123. Porto Alegre: EDIPUCRS.

CRISTÓFARO-SILVA, THAÏS. 1996. *Fonética e fonologia do português: roteiro de estudos e guia de exercícios* (2nd ed.). São Paulo: Contexto.

FAVERI, CLÁUDIA. 2000. *Duração das vogais orais em português*. Belo Horizonte: Federal University of Minas Gerais (UFMG), MS.

GORDON, MATTHEW. 1998. The phonetics and phonology of non-modal vowels: a cross-linguistic perspective. *Berkeley Linguistics Society* 24.93–105.

GORDON, MATTHEW, and PETER LADEFOGED. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29.383–406.

HOUAISS, ANTÔNIO; MAURO S.VILLAR and FRANCISCO M. MELLO. 2009. *Dicionário Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva.

ISKAROUS, KHALIL. 2010. Task Dynamics: Theory and Practice. *São Paulo School of Advanced Studies in Speech Dynamic, abstracts,* 50–62. São Paulo: SPSASSD.

LADEFOGED, PETER. 1971. *Preliminaries to linguistic phonetics*. Chicago: University of Chicago.

LEHISTE, ILSE. 1970. *Suprasegmentals*. Cambridge: MIT.

LEITE, CAMILA T. 2007. *Sequências de oclusiva alveolar + sibilante alveolar como um padrão inovador no português de Belo Horizonte*. Belo Horizonte, Brazil: Universidade Federal de Minas Gerais thesis.

MAJOR, ROY C. 1985. Stress and rhythm in Brazilian Portuguese. *Language* 61.259–282.

MARTIN, PHILIPPE. 2008. *Phonétique acoustique*. Paris: A. Colin.

MORAES, JOÃO. 1999. Um algoritmo para a correção/simulação da duração dos segmentos vocálicos em português. *Estudos de Prosódia*, ed. by Ester M. Scarpa, 69–84. Campinas: UNICAMP.

NAPOLEÃO, RICARDO F. 2010. Emergent Consonantal Clusters in Brazilian Portuguese. *Dinafon: São Paulo School of Advanced Studies in Speech Dynamic, abstracts,* 127–130. São Paulo: SPSASSD.

PARKINGSON, STEPHEN. 1990. Portuguese. *The Romance languages* ed. by Martin Harris and Nigel Vincent, 131–169. New York: Oxford University Press.

RAUBER, ANDRÉA S. 2008. An acoustic description of Brazilian Portuguese oral vowels. *Diacrítica, Ciências da linguagem* 22.229–238.

ROTHERMBERG, MARTIN. and JAMES J. MASHIE. 1988. Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech and Hearing Research* 31.338–351.

TEYSSIER, PAUL. 2001. *História da língua portuguesa*. São Paulo: Martins Fontes.

VIEIRA, MAURÍLIO N. 1997. *Automated Measures of Dysphonias and the Phonatory Effects of Asymmetries in the Posterior Larynx.* Edinburgh, Scotland: University of Edinburgh dissertation.

The revival of Añunnükü or Paraujano language

Andrés Mauricio Sabogal
University of New Mexico

ABSTRACT. Añunnükü is an Arawak language that may be described as sleeping. There is only one speaker left. Despite this fragile situation, many efforts are taking place in attempt to awaken the language. This paper seeks to describe factors that have caused the Añu people to be in this current situation, as well as to document how language loss is being prevented. The description and ensuing discussion on the vitality of the language is based within an ecological framework (Mühlhäusler 1996, Mufwene 2002). The information was gathered from the literature, six interviews, and from class observations in 2012.[*]

Keywords: Añu, Paraujano, Arawak, Linguistic Ecology, Language Revival

1. INTRODUCTION. The idea of reviving a sleeping language is a rather new phenomenon that has come about due to the increasing awareness regarding the speed of language loss within the last century. Zuckerman and Walsh (2011) propose that Australian Aboriginal language revival movements should practically model their efforts on the documented successes that Modern Hebrew went through. Although their proposal provides great insight about the linguistic issues in reviving a language, they also underestimate the complex non-linguistic factors that allowed the emergence of Modern Hebrew, such as nationalism. It is evident that the immense ecological differences between Israelis and Australian Aborigines in their respective experiences do not allow a simple modeling of efforts. Any model for reviving a language has to comprise parameters regarding the interdependency between linguistic and ecological factors.

Mühlhäusler (1996) explains that the metaphor of ecology in linguistics emphasizes the fact that a language is not detached from its context, or its environment. Just as biologists restore the niche of an endangered species in order to improve its vitality, linguists must address how to restore the system in which a specific language thrives. Mufwene (2002) believes that the most influential factors that cause language loss are changes in the socioeconomic ecology of a population. A new socioeconomic system may no longer support an ancestral language. Both of these authors' views agree that language external phenomena determine the fate of a language.

The Añu, or Paraujano, comprise approximately 21,000 people who identify themselves as the people of the water. They live on the coasts and islands of Maracaibo Lake and along the Limón River in Venezuela. They are one of the few surviving Arawak cultures of the Caribbean and today they firmly struggle to awaken their sleeping language. A young man, Yofri Márquez, who is in his thirties, is the only speaker alive today (Álvarez 2009:93). Despite their lack of speakers, poverty, and water pollution, the Añu people today face a very favorable sociopolitical environment that may allow the re-awakening of their language.

This paper will describe current efforts of reviving Añunnükü, as well as some factors in its ecology that have influenced and continue to influence its vitality. This description presents an additional reflection on linguistic revival efforts within a decolonization context. It seeks to be more comparable to the situation of Australian Aborigines and other populations in the

---

Americas. This ecological description is based on various linguistic, social, and political issues. The information in this paper has been collected from the literature, from six interviews held with academics and leaders involved in these efforts, as well as from Añu language class observations that took place in June 2012.[1]

2. LINGUISTIC ECOLOGY AND LANGUAGE REVIVAL. The ecology of a language refers to all factors external to language that may affect a language's vitality, either positively or negatively. Ignoring ecological factors becomes especially problematic for language revival or revitalization efforts. Fields such as sociolinguistics or historical linguistics have been exceptional by taking ecological factors into account when analyzing linguistic phenomena. However, linguistic ecology attempts to encompass broader factors by considering, for instance, how the physical environment also influences a specific speech community's vitality.

Linguistic ecology metaphorically parallels many concepts from biology that aid our understanding and description of linguistic issues. Biological metaphors in linguistics are already plenty. We speak of 'genetic' relationships between languages, syntactic 'trees' and lexical 'roots'. More recently even 'survival' or 'endangerment' are commonly used for language as if these were species. Proponents of linguistic ecology highlight some of these analogies by blurring the distinction between the metaphoric and the literal. For instance, biologists believe that stability within diverse ecosystems comes about via the adaptability and creativity of the species within it (Skutnabb-Kangas & Phillipson 2008). For humans, this adaptability and creativity have manifested in cultural and linguistic diversity. Skutnabb-Kangas and Phillipson believe that these two abilities are essential for the long-term planetary survival of all life. Therefore, 'language' may replace 'species' and result in working hypotheses that also add a moral touch to the reasons for maintaining linguistic diversity.

Within Latin America, literature regarding the ecological approach to language is almost non-existent. However, the inclusion of systemic factors regarding a culture's vitality is inherent in the views that indigenous peoples hold when conceptualizing language. The indigenous approach to language does not divide language use and transmission from its own specific socioeconomic setting or political contexts in which they exist. For instance, the history of linguistic domination and oppression is often invoked in Latin American indigenous schools in order to allow the maturing of the students' attitudes towards their heritage language (López 2008:141). In the last few decades, educational programs such as Ethnoeducation in Colombia, and Intercultural Bilingual Education[2] in most of Latin America, have emerged out of the need to autonomously determine how the transmission of indigenous language and culture are to be related to each group's specific environmental, political, social and economic concerns.

Furthermore, an extremely important concept that ties a language's vitality to its ecologies is the indivisibility between language, identity, and territory. Indigenous peoples must remain interconnected with their ancestral territory in order to ensure the continuity of their languages. Forceful displacement of a population amounts to a drastic change in their ecology. The likelihood that a new ecology may sustain their language is insignificant. Even in cases where displacement is not forceful, colonizers like the Norman French in England encountered unsupportive ecologies and had to shift to English (Mufwene 2002:5). The relocation of Native Americans in the US was tragically one of the most efficient language removal strategies in this country's history. In contrast, I believe that one of the most advantageous factors in the success of the revival of Modern Hebrew was the return to their land. The lexeme *eco*, from Greek *oikos*, means 'house'. It is this dedicated physical space that allows an identity, a culture, and a

language to flourish. In today's Latin America, if what we seek is the maintenance and strengthening of native languages, it is crucial that governments guarantee titles to ancestral land to every indigenous group that has not yet received it.

Blaming colonization for language loss seems to hold true only for one style of colonization. Mufwene (2002) categorizes three colonization styles: trade, exploitation and settlement. These three styles have typically resulted in the creation of pidgins, creoles and language shift respectively. Notice that the settlement colonization, as exemplified by the Spanish in Latin America, is the only style that leads to massive language shift. In the continental Americas, this type of colonization consisted of repopulating a territory and psychosocially engaging Europeans with native populations. As a result indigenous peoples were minoritized, integrated into the new socioeconomic order, and assimilated. Since the independence of most American countries, the new powers continued the oppression of indigenous cultures with the 'benign' intention of becoming homogenous nation-states.

Nowadays, there is a vigorous new interest within the field of linguistics to direct our efforts towards turning the tide of cultural extinction towards revitalization. This idea is like swimming upriver, and for revitalizing an American indigenous language, that river is a mighty one. On top of documentation, strengthening an endangered culture takes a considerable amount of ecological restoration, such as the assumption and redressing of historical burdens, as well resisting and changing various kinds of systems. For reviving a language the task is even larger.

Although one may find several decades-old literatures regarding the revival of Hebrew (Rabin 1963) and that of Cornish (Ellis 1974), language revival is still a very new concept. The Hebrew and Cornish experiences may be the only ones to have spent over a century working to bring their respective languages back to functional life. Although there are some 100 fluent speakers of Cornish today, and some are learning it as a mother tongue (Dorian 1994:488), this story is often significantly downplayed and overshadowed by proponents of the exceptionality of Hebrew revival. One finds repeated mentions of the Hebrew case as unique and the 'only' known successful linguistic revival to date (Nahir 1998:336).

But within the last twenty years, other revival movements set in contexts where settlement colonization caused language loss have emerged. For close to two decades, in the state of Massachusetts, the Wampanoag Language Reclamation Project led by Jessie Bair has been able to preliminarily awaken this Iroquoian language that had been sleeping for more than a century (Makepeace 2010). Similarly, in Southern Australia, Kaurna is being actively reclaimed after not being spoken for almost a century, and is so far considered the most successful revival movement in Australia, out of several that have been emerging (Amery 2004).

3. Añunnükü or Paraujano language

3.1. The people today. The people of the water, as they identify, have historically inhabited northern Zulia, Venezuela. They are one of the few surviving Arawak cultures of the Caribbean and today they struggle to awaken their sleeping language. The Añu are the ancestral residents of the Maracaibo Lake coasts, the islands that separate the lake from the Gulf of Venezuela and the Limón River lagoons and riversides. They live in palafittes, or stilt houses above the water, they are avid fishermen, boatmen, and consider themselves to have emerged from *ainmatuaree*, the place where the horizon meets the ocean (Quintero 2009).

The Añu, like many other indigenous groups in Latin America, have been experiencing a consciousness shift that has resulted in their revaluation of their identities. Centuries of discrimination had caused a denial of their ancestral identity. But today there is a newly born pride of being Añu. Several of my interviewees conveyed that this has been the biggest accomplishment of at least two decades of Añu cultural revitalization efforts. This dignity is most visible in the increasing results of self-determined responses to the last four censuses in Venezuela. Añu membership has increased significantly over the last couple of decades. In the 1982 census, 2,612 people declared themselves to be Añu. By 1992, the census gave a figure of 17,440 members. The 2001 census inconsistently reported 11,205 members. While the 2011 census documented over 21,000 Añu (all census results are from Mendoza et al. 2009). On the surface, we see an exaggerated population growth, but the growth is not as much in population as it is in dignity.

This psychosocial change partly results from major constitutional changes in Bolivarian Venezuela, where native peoples have achieved special privileges, as well as from a changing mainstream perspective, which is today more attentive to indigeneity as a major contributor to the re-definition of national identity. The latter factor could be exemplified by the fact that a highly modern bridge that is currently being built accross Maracaibo Lake has been named after the previously forgotten 17th century Añu chief Nigale.

3.2. GENETIC CLASSIFICATION. Añunnükü is a North Arawak language. Aikhenvald (2002) uses the labels 'Extreme North' or 'Caribbean' to refer to this branch of North Arawak, which includes Añunnükü, Wayuunaiki, Lokono,[3] Garífuna,[4] and the formerly spoken Taíno, Kaketío, and Shebayo. The Añu's closest linguistic relatives are their northwestern neighbors, the Wayuu or Guajiros. More than 300,000 people speak Wayuunaiki, and the linguistic similarities with Añunnükü are highly significant. Another possibly very closely related language could have been the one spoken by their former eastern neighbors: the Kaketío. But their language ceased to be spoken early in the colonial era and its documentation never occurred. After Wayuunaiki, the next most similar language is Lokono, although this language is spoken further east on coasts of the Guayanas. Lastly, the limited documentation of Taíno also reveals a very close relationship to Añunnükü: at least more alike than Garífuna.

Although it is linguistically obvious that Añu people should make use of the vitality of Wayuunaiki for their efforts to reawaken their own language, it is important to state that historical and political Wayuu Añu relations have today manifested an emphasis on non-Wayuu cultural features within Añu identity (Patte 1990). Consequently, there is generally no desire to model Wayuunaiki features, let alone borrow lexicon. This suggests that the needed expansion of Añunnükü lexicon, possibly brought about via comparative work, must include Lokono; a population with whom there are less historical disagreements.

3.3. LANGUAGE USE. The very densely populated Maracaibo metropolitan area surrounds Añu communities. Petroleum extraction from the lake from the beginnings of the 20th century has caused the rapid urbanization of Zulia and an overwhelming shift to the cash economy. The ethnologist, Jahn (1927:70) gathered from his observations in the first decades of the century that 'many Añu had mixed with outsiders, but even the non-Añu were speakers of their language'. The French linguist Patte (1989) found only 13 speakers in the 1970's. Patte's recognition of the rapid disappearance of Añunnükü caused her to copiously document the language during the

1970's and 1980's. She was personally acquainted with seven elder speakers in Sinamaica lagoon.

Today there is only one speaker of Añunnükü. Yofri Márquez is in his thirties, and he learned his ancestral language from his recently deceased grandmother, Ana Dolores Márquez (Álvarez 2009:93). There is possibly one more elder woman who is said to know the language. She is still alive but does not speak anymore due to very advanced old age. There are other semi-speakers, like Francisco Montiel, who despite his old age still retains partial knowledge of the language.

Besides Yofri Márquez's extremely rare case of intergenerational transmission, there are many Añu and non-Añu learners of the language. These learners do not use the language functionally (yet), but they represent the primary great hope for the survival of the language. Lastly, it is not superfluous to state that linguists Patte, Álvarez, Mosonyi and Bravo possess very valuable knowledge that is crucially important for the fate of the language.

The only interactional use of the language seems to be the ethnically symbolic use of words or short phrases embedded in dominant *Maracucho* Spanish. This symbolic use evidences the advanced stages of identity reclamation, as well as the incipient stages of reviving the language. Non-interactional uses of the language however seem to be more advanced. There is for instance an Añunnükü version of the Venezuelan national anthem, as well as the *Pater Noster* and the *Ave Maria* prayers. But as expected, the language emerges mostly in schools in Añu communities. Children are receiving Añunnükü instruction in three elementary schools in Santa Rosa de Agua,[5] two others on the Sinamaica lagoon,[6] and one in Maraca Island.[7] It is in these spaces indeed, where children are partially acquiring the language, and thus they constitute the hope for concrete results in this language revival effort.

In El Moján, the Kanüye Añu[8] School is unique and significant in that it trains Añu teenagers or adults to revitalize and develop their native fishing economy, traditional arts and crafts, as well as the language. Additionally the Misión Sucre in Santa Rosa de Agua gives Añu language and culture instruction for adults. Nevertheless, a great obstacle towards revival is that none of the instructors of the language are fluent, and therefore, apprentices acquire basic knowledge of the language at best. The only exception is when Yofri Márquez teaches for special occasions.[9] Overall, it is clear that the language is not alive, but it is also evident that the efforts to reverse its disappearance are momentous and praiseworthy.

3.4. AÑUNNÜKÜ DOCUMENTATION. In the colonial documents there is mention of the people, but there seems to be almost no descriptions of the language. In the 20th century the first vocabulary lists and minor grammatical items were produced. Jahn (1914, 1927), Oramas (1918) and Rivet and Wavrin (1952) published mainly short vocabulary lists. This was practically all the documentation that the language had by the time that Marie-France Patte began her Añunnükü work. Patte (1978, 1981, 1986, 1987, 1990, 1992) published various linguistic and anthropological articles as well as her invaluable grammar titled *Estudio Descriptivo de la lengua Añun* (1989).

Around the turn of the century, Venezuelan linguists began collaborating with the communities and with UNICEF to compile all existing documentation and to create pedagogical materials that would allow the language to be taught in Añu schools to all children. They have been able to compile the dialogue book *Anii Wanükü* (Álvarez 2007) and a basic dictionary containing 3,000 words, *Liuruchaa Añunnükümoyatü* (Álvarez & Bravo 2008). Álvarez (2009) then continued with the publication of an article on syntax and wrote a grammatical sketch,

Esbozo gramátical de la Lengua Añu, which is yet to be published. Additionally, Yofri Márquez collaborated with Álvarez, Bravo, and Fernández (among others), in turning a traditional legend told by his grandmother Ana Dolores Márquez into a children's book with an accompanying CD titled *Pütümaata keichi Pütümaata 'sleep moon, sleep'* (Márquez 2007).

These latest and most important writings supporting language revival (Álvarez 2007, Álvarez & Bravo 2008, Álvarez 2009, Márquez 2007) are written following the Venezuelan Indigenous Languages Alphabet conventions. This has defined the orthography in a very practical manner. A significant change between Patte's (1989) and Álvarez's (2007) proposed orthographies is a simplification of the vowel system. Patte described the existence of two high central vowels: unrounded [ɨ] and rounded [ʉ], and suggested the graphemes <ï, ü> respectively. But Álvarez believes that [ʉ] is an allophone of [ɨ] and thus this phoneme should be written as <ü>, as in Wayuunaiki. Furthermore, Patte described the existence of a mid central rounded vowel [ɵ], written as <ö>. But she also writes that both [ɵ] and [ʉ] are infrequent phonemes that only occur in vowel geminates (Patte 1989:21), and that [ɵ] is the only vowel that does not participate in diphthongs (Patte 1989:27). Álvarez interprets the available data and Patte's descriptions of the limited distribution of these rounded central vowels as indicative that these were likely allophones. This interpretation coupled with the complexity of transmitting three distinct central vowels to native Spanish speakers without any intergenerational transmission, led to the decision to replace the rare occurrences of <ö> with either /e/ or /o/, as in Wayuunaiki (Álvarez 2007:21). In sum, the works of Patte and Álvarez definitely constitute the most important linguistic written documentation of the language to date.

But a more powerful form of documentation for the purpose of transmission has also taken place. In 2005, the Venezuelan ministry of culture presented a one-hour video documentary, *Somos Añu*. This video is an introduction of the people and their culture to Venezuelans, but it also presents their language revival efforts. In another video production with much greater revitalizing potential, Matheus and Matheus (2011) created a documentary series to be broadcast on national television, called *Los Añú: Una Lengua Patrimonial en Recuperación (The Añu: A recovering heritage language)*. This series consists of eleven 30-minute episodes describing their culture and language and their monumental efforts to reestablish themselves as Añu. These five-and-a-half hours of high quality audiovisual documentation features Yofri Márquez as a narrator and translator. Roughly half of the material is spoken in Añunnükü with subtitles in Spanish, while the other half is in Spanish with Añunnükü subtitles. Also in 2011, Añu filmmaker from Santa Rosa, Gretzy Atencio produced another TV series titled *Los Niños del Mangle (The Mangrove children)*, which is a more creative production that specifically celebrates Añu culture, although it is mostly in Spanish.

3.5. Institutional support. When it comes to language, the efficiency of policies may be questionable. But if Añu language revival requires its ecologies to provide the necessary space and support, the law on their side definitely fuels their process. In 1999, Venezuela reformed its Constitution. This new constitution, following the trend of constitutional reforms in the region, conceded many rights and benefits to indigenous Venezuelans. Among these, indigenous languages have become official in their own communities (Art. 9).

Although Article 119 guarantees their right to their ancestral territory, both Añu and Wayuu are still waiting for land titles. The environment of the Añu area has also been negatively affected for many decades by oil extraction generated pollution of their waters. Now Article 120 maintains that the socioeconomic or cultural integrity of the peoples is not to be harmed by

resource extraction. Sadly, the water pollution in the lake is severe, and given that the Añu economy is based around seafood, their integrity has clearly been harmed.

In the field of education, native Venezuelans have the freedom to implement their own curriculum within Intercultural Bilingual Education (Art. 121). This means that the classes must be taught in indigenous languages, as well as in Spanish, so that bilingualism and the ability to communicate both within and outside of one's culture becomes the norm for indigenous graduates. This promises to be a very long project. In most of Latin America this educational philosophy only exists in elementary schools. As for the Añu, their students must legally learn their language, but they lack speakers to teach it.

Furthermore, in 2009, the National Assembly of Venezuela passed the Indigenous Peoples Cultural Patrimony Law. This very recent law, among other mandates, prohibits copyrighting indigenous cultural heritage. Additionally, this law guarantees the 'transmission, diffusion and revaluation' of the cultural manifestations of communities that face 'the threat of extinction' (Art. 17). This is to be realized by compiling and safeguarding such manifestations in literary or audiovisual form, or by using new technologies. This same article also suggests that indigenous people be the protagonists and beneficiaries of the necessary documentation and research.

Outside of Venezuela, UNICEF has also been engaged, primarily with the Sinamaica communities, by contributing financial support within the last decade or so for the purpose saving their language and culture. They have already built a school, in the traditional palafitte style, and contributed financial aid for the publication of a dictionary (Álvarez & Bravo 2008) and a dialog book (Álvarez 2007).

4. DISCUSSION. This paper has attempted to present a picture of Añu language vitality by briefly exploring some linguistic and non-linguistic factors. On the linguistic side, much has been done with very limited resources. The dictionary (Álvarez & Bravo 2008) only has 3,000 words. It is estimated that there were around 7,000 words documented in Hebrew before they began their revival process of word creation (Rabin 1963). This is an area of concern and students of the language should know that the only two ways to increase the vocabulary is by either inventing words or borrowing them. Given the rather tense Wayuu-Añu dynamics, it is important to also consider Lokono and Garífuna lexicons when proposing neologisms.

Eira and Couzens (2010) who have been very active in the revival of Australian Aboriginal languages suggest that the lexical and grammatical processes described in language contact and language change theories are highly relevant to reviving languages. For instance, they propose the idea of language mixing in reverse. Here, the dominant language to which a culture has shifted should be gradually 'infiltrated'. The specific suggested order would first focus on the use of ancestral lexicon, with some attention to its phonological features. Then native syntactic features are to be used in the dominant language, and lastly the native morphology.

These authors believe that revival must occur through mixed language phases. Zuckerman and Walsh (2011) for instance, posit that Modern Hebrew is a Semito-European hybrid. They claim that the language is primarily a Hebrew Yiddish mix with elements from many other languages (Zuckerman and Walsh 2011:114). Thus language revival also involves its evolution. As we know languages change with time, and those which stop being spoken for a period of time may undergo greater change. For that reason it is imperative that linguistic purism does not hinder revival work. It makes sense that speaking a language with limited resources will

generate major linguistic gaps that must be filled on the spot. Communication then takes precedence as the major practice that will revive the language, in which there will be plenty of need for dominant language lexicon and structure in order to ensure success. As for the Añu, the currently limited knowledge that apprentices have acquired should be used in communication with help from the Spanish language. This is phase one of revival.

On the ecological side there is considerable legislative and institutional progress. The Añu are building schools designated to implement mandatory bilingual education. The constitution and the cultural patrimony law guarantee significant cultural rights, benefits and protections. As for the environment, there is one huge problem: water. Alí Fernández, a Wayuu social anthropologist who is one of the most important figures in the decades-long reclamation of Añu identity stated in an interview that I conducted, 'the identity of the millenarian people of the water, has been heavily impacted by their acknowledgment that for the first time in history they are the people of the polluted water.' The new constitution guarantees that oil extraction in the Maracaibo basin must not affect Añu socioeconomic or cultural integrity (Art. 120). Contaminated waters not only affect Añu culture and economy, but also jeopardize their actual future existence.

5. CONCLUSION. Impressive and significant linguistic and ecological progress has provided a very important stepping stone towards the revival Añunnükü. Apprentices must be aware that even if they attempt to learn the 'pure' language, a successful revival will result in a new variety, which could be thought of as Neo-Añunnükü. The vocabulary needs to be expanded based on comparative work. It is highly important that Lokono and Garífuna are included in these comparisons. Lastly, I agree with Patte (2011) that the concurrent processes of purifying the waters and receiving land titles constitute the two most important ecological factors for success in linguistic revival.

REFERENCES

AIKHENVALD, ALEXANDRA. 2002. *Language contact in Amazonia*. New York: Oxford University Press.

ÁLVAREZ, JOSÉ, and MARÍA BRAVO. 2008. *Diccionario básico de la lengua añú. Liuruchaa añunnükümoyatü*. Maracaibo: UNICEF-Universidad del Zulia.

ÁLVAREZ, JOSÉ. 2007. *Anii wanükü. Aquí está nuestra palabra*. Maracaibo: UNICEF-Universidad del Zulia.

ÁLVAREZ, JOSÉ. 2009. La oración simple y la oración compuesta en añu. *Lingua Americana* 13.91–140.

AMERY, ROB. 2004. Kaurna language reclamation and the formulaic method. *Language is Life: Proceedings of the 11th Annual Stabilizing Indigenous Languages Conference, University of California, Berkeley* 81–99.

ATENCIO, GRETZY (producer, director). 2011. *Los Niños del Mangle* [Television broadcast]. Zulia, Venezuela (available from producer).

DORIAN, NANCY. 1994. Purism vs. compromise in language revitalization and language revival. *Language in Society* 23.479–494.

EIRA, CHRISTINA, and VICKY COUZENS. 2010. Meeting point: Parameters for a typology of revival languages. *Victorian Aboriginal Corporation for Languages document*. Online: http://www.vaclang.org.au/project-detail.aspx?ID=28.

ELLIS, PETER. 1974. *The Cornish language and its literature*. London: Routledge & Kegan Paul.

JAHN, ALFREDO. 1914. *Parauhanos und Guajiros und die Pfahlbauten am see von Maracaibo*. Berlin: Zeitschrift für Ethnologie.

JAHN, ALFREDO.1927. *Los aborígenes del occidente de venezuela. Su historia, etnografía y afinidades lingüísticas*. Caracas: Litografía y Tipografía del Comercio.

JEROZOLIMSKI, DANIEL (producer), and ALEJANDRA SZEPLAKI (director). 2005. *Somos Añu* [Television broadcast]. Venezuela: Cooperativa Estrella Films.

LEY DE PATRIMONIO CULTURAL DE LOS PUEBLOS Y COMUNIDADES INDÍGENAS. 2009. *Gaceta oficial de la República Bolivariana de Venezuela* No. 39115.

LÓPEZ, LUIS. 2008. Indigenous contributions to an ecology of language learning in Latin America. *Ecology of language*, ed. by Angela Creese, Peter Martin and Nancy Hornberger, 141–55. New York: Springer.

MAKEPEACE, ANNE (director). 2010. *We Still Live Here: Âs Nutayuneân* [DVD]. United States: Makepeace productions.

MÁRQUEZ, ANA DOLORES. 2007. *Pütümaata, keichi, Pütümaata ¡Luna, luna, no te duermas!* Caracas: Monte Ávila.

MATHEUS, GILBERTO (producer, director, writer); MATHEUS MARÍA (producer); ALÍ FERNÁNDEZ (writer); and GILBERTO POLO (writer). 2011. *Los Añú. Una Lengua en recuperación.* [Television broadcast]. Zulia, Venezuela (available from producer).

MENDOZA, ILEANA; ELIZABETH SEGOVIA; and ALÍ FERNÁNDEZ. 2009. Tendencias de las políticas de transición en las comunidades rurales, indígenas y de frontera. Estudio de caso: Venezuela. *Organization of American States report*. Online: http://portal.oas.org/Portal/Topic/SEDI/Educaci%C3%B3nyCultura/Educaci%C3%B3nCuidadoyDesarrolloInfantil/Proyectos/Transicionesexitosasdelni%C3%B1o/tabid/1317/Default.aspx.

MÜHLHAÜSLER, PETER. 1996. *Linguistic ecology: language change and linguistic imperialism in the pacific rim*. New York: Routledge.

MUFWENE, SALIKOKO. 2002. Colonisation, globalisation, and the future of languages in the twenty-first century. *International Journal of Multicultural studies* 4.162–93. UNESCO journal. Online: http://www.unesco.org/new/en/social-and-human-sciences/resources/periodicals/diversities/past-issues/vol-4-no-2-2002/colonisation-globalisation-and-the-future-of-languages-in-the-twenty-first-century/.

NAHIR, MOSHE. 1998. Micro language planning and the revival of Hebrew: A schematic Framework. *Language in Society* 27.335–357.

ORAMAS, LUIS. 1918. Patronímicos quiriquires y vocabulario paraujano comparado con el Guajiro. *De Re Indica* 1.43.

PATTE, MARIE-FRANCE. 1978. Etude phonologique de la langue añun. Parlee dans la région de Sinamaica (Vénézuéla). *Amerindia* 3.57–83.

PATTE, MARIE-FRANCE. 1981. Les préfixes personnels en añun - Morphophonologie. *Amerindia* 6.7–16.

PATTE, MARIE-FRANCE. 1986. *De los añun*. Paris: Chantiers Amerindia.

PATTE, MARIE-FRANCE. 1989. *Estudio descriptivo de la lengua añun o paraujano*. San Cristóbal (Venezuela): Univ. Católica del Táchira.

PATTE, MARIE-FRANCE. 1990. Entre guajiros et espagnols, les humains récits añun. *Bulletin de l'IFEA* 19. 499–506.

PATTE, MARIE-FRANCE. 1992. Variations d'actance en añun. *Actances* 3.178–94.

PATTE, MARIE-FRANCE. 2011. Territory, identity, and language among the Añun people. *Ethnographic Contributions to the Study of Endangered languages*, ed. by Tanisa Granadillo and Heidi Orcutt-Gachiri. Tucson: The University of Arizona Press.

QUINTERO, JOSÉ. 2010. Tradición oral y construcción de la historia. *Revista Cifra Nueva* 21.23–38.

RABIN, CHAIM. 1963. The revival of Hebrew as a spoken language. *Journal of Educational Sociology* 36.388–392.

RIVET, PAUL, and ROBERT DE WAVRIN. 1952. Les Indiens Parawgwan. *Journal de la Société des Américanistes* 41.235–238

SKUTNABB-KANGAS TOVE, and ROBERT PHILLIPSON. 2008. A human rights perspective on language ecology. *Ecology of language*, ed. by Angela Creese, Peter Martin and Nancy Hornberger, 3–13 . New York: Springer.

SUAZO, SALVADOR. 2011. *Lila Garifuna. Diccionario garífuna-garífuna-español*. Tegucigalpa: Litografía López.

ZUCKERMAN, GHIL'AD, and MICHAEL WALSH. 2011. Stop, revive, survive: Lessons from the Hebrew Revival Applicable to the Reclamation, Maintenance and Empowerment of Aboriginal Languages and Cultures. *Australian Journal of Linguistics* 31.111–127

NOTES

[1] IRB Protocol #12–137. The Awakening of the Añu Language and its Continental Implications

[2] Bolivia, Peru, Ecuador, Colombia, Venezuela, Paraguay, Nicaragua, Mexico and Guatemala are among the countries that have passed legislation establishing mandatory Intercultural Bilingual Education (EIB) for regions where indigenous peoples reside. Although the full implementation of this evolving educational philosophy has proven to be a major challenge to most countries, the evolution of the model has aroused major intercontinental debates in which the tenets of linguistic ecology are always on the table.

[3] Lokono, also called Arawak, is spoken in French Guiana, Surinam, Guyana and the Venezuelan Guayana. There are just over 2,000 speakers in these four countries.

[4] Garífuna has been previously called Black Carib, and it is effectively Modern Island Carib, also called Kalíphuna or Eyeri. Island Carib was formerly spoken in the lesser Antilles, primarily in the islands of Yurumain (Saint Vincent), Galugaira (Guadaloupe) and Wai Tugubuli (Dominica) The speakers from Yurumain maintained the language in their new lands of Central America. After the British deportation from the island, they established themselves in the coastal lands of Belice, Guatemala, Honduras and Nicaragua. Today there are more than 150,000 speakers.

[5] Batalla de Carabobo, Batalla de Boyacá and Jesús María Portillo

[6] Bartolomé Duarte and Laguna Sinamaica

[7] Maraca

[8] Escuela Técnica Agropecuaria Kanüye Añu

[9] The state of Zulia has not granted teaching credentials to Yofri.

Blackfoot demonstratives in narrative discourse: A pragmatic analysis

S. Scott Schupbach
University of Montana

ABSTRACT. Previous analyses of the Blackfoot demonstrative system (Uhlenbeck 1938, Taylor 1969, Frantz 2009) focus on situational uses of demonstratives and often conflate the meanings of multiple pragmatic uses of demonstratives into overly broad definitions. In this paper, I discuss the meanings of Blackfoot demonstratives in terms of their pragmatic functions. I follow Himmelmann's (1996) taxonomy of universal demonstrative uses as my criteria for delineating demonstratives into four categories: situational; discourse deixis; tracking; and recognitional. I demonstrate that recognizing the different pragmatic functions of demonstratives allows for a more nuanced description of their meanings and usage.

Keywords: Blackfoot language, demonstratives, anaphora, discourse deixis, recognitional demonstratives, pragmatic uses of demonstratives

1. INTRODUCTION. Cross-linguistically demonstratives serve a number of pragmatic functions. While the core meaning of directing the attention of the addressee to a particular entity is encoded in each of these pragmatic functions, the type of referent, the addressee's familiarity with the referent, and the knowledge shared by the speaker and addressee are among the factors that vary across the different pragmatic uses of demonstratives. In this paper, I discuss the pragmatic uses of demonstratives found in Blackfoot narrative discourse. In doing so, I provide evidence in support of Himmelmann's (1996) claim that demonstratives serve four universal pragmatic functions and illustrate the need for describing non-situational pragmatic functions of demonstratives in descriptive grammars.

In §2, I provide a brief overview of recent typological work on pragmatic functions of demonstratives. In §3, I provide data and analysis from Blackfoot textual data. The texts I analyze are from a collection of stories, mostly short stories, elicited and transcribed by Uhlenbeck (1912). The collection contains both traditional etiological stories, as well as historical and first-person accounts of Blackfeet life and customs. In §4, I provide some concluding remarks and discuss issues for further research.

2. TYPOLOGICAL GENERALIZATIONS OF DEMONSTRATIVE USES. Himmelmann (1996) and Diessel (1999) divide pragmatic functions into two main categories. The first is termed SITUATIONAL as this type is used to refer to non-linguistic entities situated in the physical environment. This type is sometimes also called exophoric as the referent is outside of the discourse itself. The second type is known as NON-SITUATIONAL, or endophoric, and its uses encode referents that are part of the discourse event. The non-situational category is further delineated into three sub-groups: tracking (or anaphoric) uses; discourse deixis; and recognitional uses.

Himmelmann (1996) proposes that situational, tracking, discourse deictic, and recognitional uses are universal. He argues further that because they are each universal, they are equally basic, that is, none of the pragmatic functions is derived from another. Although these four pragmatic uses are attested in much of the available cross-linguistic data on demonstratives, that data is severely limited by the fact that most descriptive grammars do not address non-situational pragmatic roles of demonstratives. Because of this, Diessel (1999) argues that

Himmelmann's universality claim is based on too little data and that further research is needed. Diessel (1999) also argues against Himmelmann's basic use claim, stating that situational uses are morphologically less complex, syntactically less restricted, and developmentally first-acquired and are therefore the basic demonstrative use from which non-situational uses are derived. I return to this second claim in §4.

2.1. SITUATIONAL USES. Situational uses of demonstratives may be broadly defined as those that point out non-linguistic entities in the physical environment (Diessel 1999:6). Diessel (1999), following Fillmore 1997, categorizes situational uses into gestural and symbolic uses, both of which encode features of spatial demarcation, but which differ with respect to whether entities are locatable in the surrounding environment (gestural) or in an imagined physical space (symbolic). Symbolic demonstrative uses often refer to entities in the surrounding environment that are either (i) too large in scope to be completely visible, or (ii) are non-corporeal. Symbolic demonstratives may also be used by the speaker to mentally situate the addressee within the physical environment of a narrative setting. This subtype of symbolic use is referred to by Lyons (1977) as deictic projection. In discussions of deictic projection in this study, I distinguish between the setting of the narrative event and that of the narration event. I use the term NARRATIVE SETTING to refer to the location where the story takes place as described by the narrator. NARRATION SETTING, on the other hand, refers to the location where the speech act participants are physically located when the speaker relays the story to his audience.

2.2. NON-SITUATIONAL USES. In addition to directing the addressee's attention to entities in the physical environment (or a mental representation of a physical environment), demonstratives also serve to direct the addressee's attention to participants that recur throughout the discourse. These uses of demonstratives interact with other tracking devices referred to as anaphora (e.g. personal pronouns, definite and indefinite articles, etc.). Because the term anaphoric is applicable in other broader contexts, Himmelmann (1996) prefers the term TRACKING to describe this pragmatic use of demonstratives. Himmelmann (1996) states that compared to other anaphoric devices, tracking demonstratives are relatively infrequent, and that they are used to reference major discourse participants. Diessel (1999:96) suggests that anaphoric demonstratives are often used to 'indicate a referent that is somewhat unexpected and not currently in the focus of attention'.

Like tracking usage, in discourse deixis usage, demonstratives refer to entities within the discourse itself. However discourse deictic uses differ from tracking uses in that the former do not refer to specific NPs, but to larger chunks of linguistic content. Discourse deixis focuses on 'aspects of meaning expressed by a clause, a sentence, a paragraph, or an entire story' (Diessel 1999:101). Alternately, tracking usage focuses on participants or objects within the narrative. The primary function of discourse deixis is to link the proposition in which the demonstrative is embedded to the proposition to which the demonstrative refers. Diessel (1999) states that discourse deixis is both anaphoric (backward-referring) and cataphoric (forward-referring). Cataphoric reference is described as one of the features that distinguishes the anaphoric pragmatic category from the discourse deixis category (Diessel 1999, Levinson 1983), but how common cataphoric reference is cross-linguistically requires further examination. The lack of typological generalizations regarding cataphoric reference is due in large part to the absence of descriptions of pragmatic demonstrative uses in descriptive grammars. This reveals the need for further research into discourse deictic uses in individual languages.

Although the recognitional use of demonstratives is first proposed in Lakoff 1974 and referenced in a number of subsequent studies, Himmelmann (1996) provides the first detailed discussion and is the first to argue that the recognitional usage is a universal pragmatic function of demonstratives. Recognitional uses have four defining features: they only occur adnominally; they are used to activate specific knowledge that is shared by the speaker and addressee; they refer to information that has not yet been mentioned in the preceding discourse; they refer to entities that are not locatable in the surrounding environment (Diessel 1999). That is, they are adnominal, non-situational demonstratives and their referents are hearer-old, but discourse-new.

2.3. THE NEED FOR PRAGMATIC ANALYSIS IN DESCRIPTIVE GRAMMARS. Most descriptive grammars categorize demonstratives by morphosyntactic characteristics. Descriptions often include paradigms of pronominal, adnominal, adverbial, and/or identificational forms, however glosses nearly always reflect situational meanings alone, or overly broad meanings that represent conflations of situational and non-situational uses (Himmelmann 1996). Examining demonstratives based on syntactic functions is not sufficient for describing the full range of demonstrative uses and meanings. I illustrate this fact by comparing 1 through 4 below. In 1, the referent is not overtly stated by the speaker, but is indicated by gesture, and the demonstrative stands alone as a pronoun. In 2, the referent is overtly mentioned along with an adnominal demonstrative.

(1) Look at **that**! (Accompanied by a gesture toward a bird.)
(2) Look at that bird! (Accompanied by a gesture toward a bird.)

Both of these exclamations may be made about the same referent, for instance, a bird that is hanging upside down like a bat from a telephone wire. In both examples, 'that' is functioning in the same way, to direct the addressee's attention to a bird and to indicate that the bird is not near the speaker. In these two cases, one set of semantic features is sufficient to describe both occurrences of 'that' even though the word is functioning pronominally in 1 and adnominally in 2. Now compare 3 and 4, both of which contain adnominal demonstratives.

(3) I bought that car this morning. (accompanied by a gesture toward a new car)
(4) I bought **that** car this morning. (no accompanying gesture)

In 3, there is a specific referent visible to the addressee toward which the speaker makes a gesture. In this case, the meaning of 'that' is the same as it is in 1 and 2, (i.e. it is being used to point out a referent that is in the surrounding environment but that is not near the speaker). However, in 4 without an accompanying gesture, the assumption being made by the speaker is that the addressee knows which car the speaker was planning to purchase even though it is not present in the physical environment and has not been previously introduced in the current dialogue. The use of 'that' in 4 does not indicate anything about the spatial demarcation of the referent. This is the recognitional use described in §2.2 above. The meaning of 'that' in 4 indicates shared knowledge between the speech act participants, so spatial semantic features are of no use in assigning meaning to 'that' as it occurs in this example.

Of these four examples, 2 - 4 are identical morphologically and syntactically but not pragmatically or semantically. On the other hand 1 - 3 pattern together morphologically, pragmatically, and semantically, though not syntactically. A syntactic taxonomy is often of great

value in studies of demonstratives and is appropriate for examinations of morphosyntactic form, especially in languages where nominal, adverbial, verbal, and identificational forms vary morphophonologically. However, as seen here, for the purpose of investigating semantic content, a pragmatic taxonomy is required as pragmatic function plays a large role in determining meaning. Ultimately, both are useful distinctions for an in-depth analysis as each taxonomy highlights different aspects of function, form, and meaning.

3. BLACKFOOT DEMONSTRATIVES. In this section I describe the pragmatic uses attested in the Blackfoot stories transcribed in Uhlenbeck 1912. Here I provide evidence in support of Himmelmann's (1996) claim that all four pragmatic uses are universal by demonstrating the presence of all four categories in Blackfoot.

3.1. SITUATIONAL USES. In Schupbach 2013 I describe gestural situational uses of Blackfoot demonstratives, summarized in Table 1 below.

Stem	Geometric Configuration	Number/ Gender	Referent/Region Configuration
am 'proximal with respect to the speaker'	*-o* 'interior'	*-wa* 'proximate singular'	*-ma* 'stationary'
ann 'medial with respect to the speaker'		*-yi* 'obviative singular'	*-ya* 'moving away from speaker' (*-ia* in Uhlenbeck)
		-yi 'inanimate singular'	
om 'distal with respect to the speaker'		*-iksi* 'animate plural'	*-ka* 'moving toward speaker'
		-istsi 'inanimate plural'	*-hka* 'invisible, indiscernible'

TABLE 1. Blackfoot demonstrative template (partial).

In Schupbach 2013 I argue that in gestural situational uses the demonstrative stems *am, ann,* and *om* encode the anchor feature [speaker] and the spatial demarcation features [proximal], [medial], and [distal] respectively. The suffix *-o* is used when the referent is located in a space shared by the speaker and the addressee − Imai's (2003) [interior] geometric configuration parameter. This suffix only occurs in situational uses. The suffixes *-ma, -ya, -ka,* and *-hka* encode spatial deixis features of referent/region configuration, which indicate features related to the motion or invisibility of the referent. The suffix *-ma* 'stationary' is frequently used to refer to stationary objects, the position of which in relation to the speaker is determined by the stem. The suffix *-ya* 'motion away from speaker' is used when the referent is moving away from the speaker, and is also used in offerative cases when the speaker is handing the referent to the addressee. The morpheme *-ka* 'motion toward the speaker' is used with referents moving toward the speaker, but also has the extended spatial meaning of 'back' and the extended temporal meaning of 'past'. The suffix *-hka* 'invisible, indiscernible' refers to objects that are in the surrounding environment but are concealed, or that have just left the area and are just out of sight. As these gestural situational uses are addressed in detail in Schupbach 2013, the remainder of the current study focuses on symbolic situational uses and non-situational uses.

In the textual data, Uhlenbeck occasionally includes contextual notes enclosed in brackets that help the reader understand the setting and gestural information conveyed by the storyteller. But when such information is not provided, the present-day reader knows little about the audience or the circumstances of the narration event of the stories given in Uhlenbeck 1912. Thus it is often difficult to determine exactly how symbolic situational demonstratives function in the textual data. For example, in 5 below it is possible that the storyteller is using the proximal demonstrative form either (i) as a means of situating himself and his audience within the narrative setting (deictic projection) or (ii) because he is physically near the river discussed in the story and is equating the narrative setting with the setting of the narration event (wider-context).

(5) Ksiskɑniáutunìi   ototoåχkàni,
     ksisskanáótonni-yi   oto-otowoohkaa-n-yi
     morning-IN.SG[1]   go-find.horse.AI-NOM-IN.SG
         'In the morning, when he went for the horses [to bring them into camp] …'

     amóia   niétɑχtaii   áitòtò.
     am-o-ya   niítahtaa-yi á-it-o'too-wa
     DP-INT-MA   river-IN.SG   DUR-DCT-arrive-3SG
         'he came to **this** river.'   Uhlenbeck (1912:65)

That the river is mentioned by name earlier in the story suggests that deictic projection is a more likely explanation. If the river were close enough that it could be referred to by means of a gestural or wider-context use, then it is likely that that strategy would be used to introduce the river into the narrative. Using the river's name in the first mention suggests that the river is one whose name is recognizable, but which is not near enough to gesture toward or to be considered part of the wider-context identified by the *am* stem. Thus, it is most likely that the deictic projection use of the proximal stem to refer to the river in 5 is a narrative strategy employed to make the audience feel a part of the story.

        In 6, there is a similar example in which a lake is referred to with the proximal stem *am*. But this time, the proximal stem is used on the first mention suggesting a wider-context use, rather than a deictic projection use.

(6) **Amóm**   ómaχksikimìm   kináutamisò
     am-o-ma   omahk-ikim-yi-ma   ikkina-á-ot-waamis-oo-wa
     DP-INT-STAT big-water-IN.SG-STAT slow-DUR-go.do[2]-up-go.AI-3SG
         'He went slowly up to a lake **in this country here**.'   Uhlenbeck (1912:180)
         Literally: 'He went slowly up to **this** lake.'

The presence of the *am* stem at the first mention of the lake and the translation of *amóm* with the phrase 'in this country here' both suggest a wider-context symbolic use. Because this is the first sentence of the story and no lake or setting has yet been described, the function of *amóm* must be to ground the story in the setting of the narration event (wider-context) rather than to bring the audience into the setting of the narrative (deictic projection). The use of the phrase 'in this country here' in the English translation is likely intended to capture this pragmatic function of the demonstrative. Additionally, since many of the narratives in Uhlenbeck's collection speak to the Blackfeet people's connection to the land and its resources (especially the buffalo herds), it is

135

very likely that many of these stories are set in a region to which the storyteller feels he belongs. Thus, I analyze non-gestural uses of the *am* stem, which refer to geographical features without an overtly expressed setting, as wider-context symbolic uses. Conversely, when the *am* stem is used with geographical referents in stories where the setting has been explicitly described (usually by means of place names), I analyze these as deictic projection symbolic uses, as is the case in 5 above.

3.2. TRACKING USES. Tracking uses are the most common in textual data I analyzed from Uhlenbeck 1912 and comprise almost two-thirds of 135 demonstrative tokens. About 61% of these tracking uses contain the stem *om* and the other 39% are built from the *ann* stem. The *am* stem does not function anaphorically.

The distal stem *om* is often used to introduce new information. Although I found no statistically significant correlation between *om* and discourse participants that are the current topic or focus of their clause,[3] further study may show that it is a component of Blackfoot's focalization system. Introductory-*om* functions similarly to introductory-*this* in English in that its referents seem to be more likely to recur throughout the narrative.[4] An example of introductory-*om* is given in 7.

(7) **Omík**   kitsímik   káuaiχtsiu.
    om-yi-ka   kitsímm-yi-ka ikawaihtsi-wa
    DD-IN.SG-MT door-IN.SG-MT be.open.II-3SG
       '**The** door lay open.'   Uhlenbeck (1912:65)

Before the sentence in 7, no mention of a door has yet been made. While the setting of this section is Four-bears' lodge, and it could be considered common knowledge that there is a door, this is the first overt mention of it in the story. It is likely that the door is overtly mentioned here because its state of being open is relevant for the next few clauses.

The stem *om* is also used to refer to previously mentioned discourse participants that are temporarily relevant, but have not recently been in focus. An example of this is given in 8 in which *omá* refers to a man, Four-bears, who was not an argument in the previous three clauses, but who was introduced into the narrative earlier.

(8) **Omá**   Nisóχkyaio   áisαmò   itsiksískaχkuyìuaie.
    om-wa niisó-kiááyo-wa   á-isamo-wa   it-iksisskahko-yii-wa-áyi
    DD-3SG four-bear-3SG   DUR-be.a.long.time.II-3SG DCT-nudge.TA-DIR-3SG-DTP
       'After a long while, Four-bears pushed it.'   Uhlenbeck (1912:65)

Omá serves to temporarily emphasize Four-bears, even though the bird is still the main character of this portion of the narrative. In fact, after this clause the bird is the only actor for the remainder of the story.

In Blackfoot, anaphoric uses of *om* bear much of the tracking load shared by pronouns and definite articles in other languages. Thus *om* is often translated with English articles *the* and *a(n)*, and sometimes with the English distal demonstrative *that*. However unlike English determiners and demonstratives, Blackfoot demonstratives do not encode definiteness or specificity (Glougie 2000, Genee 2005).

The medial stem *ann* is never used to introduce new participants but only refers to previously given information that has been recently mentioned. When the *ann* stem is used, the first mention of the referent is often in the same sentence. Occasionally, the tracking use of the demonstrative links a nominal in the left-detached position to a clause, a position that often contains focal or topical arguments (see Van Valin & LaPolla 1997 for discussion of the significance and use of this extra-clausal position). In the textual data, when the *ann* form and its referent are not in the same sentence, the referent is in the sentence immediately preceding. An example of this use of *ann* is presented below in 9.

(9) Akimóχtsim   **ánnimaie**   itáupiu   omá   pìksíu.
     akim-ohtsi-ma   **ann-yi-ma-ayi**   it-á-opii-wa   om-wa   pi'ksii-wa
     place.of.honor-LOC-STAT **DM-IN.SG-STAT-ID** DCT-DUR-sit.AI-3SG DD-3SG   bird-3SG
          'The bird sat at **the** upper end of the lodge.'   Uhlenbeck (1912:65)
          Literal translation: 'The place of honor, **that is** where the bird sat.'

In 9, the anaphoric demonstrative refers to information that was just mentioned the word before in a left-detached position. The emphasis on where the bird sat is relevant to the setting and the later visual imagery of the bird flying across the lodge and out through the door.

3.3. DISCOURSE DEICTIC USES. In Blackfoot, only the *ann* stem is used for discourse deixis. The function of discourse deixis – to link the proposition in which the demonstrative is embedded to the proposition to which the demonstrative refers (Diessel 1999:101-102) – is illustrated in 10.

(10) **Kɛ́nnyaie**   mátanistsinoàu   ksistsikúma.
     ki-ann-yi-ayi   matt-aanist-ino-aa-wa   ksiistsikomm-wa
     and-DM-IN.SG-ID   also-MNR-see.TA-DIR-3SG   thunder-3SG
          '**That way** the thunder was seen.'   Uhlenbeck (1912:65)

In this example, the discourse deictic use of *kénnyaie* refers to the preceding five sentences of the story, which describe how the bird was responsible both for the flashes of lightning as well as the sound of thunder. The demonstrative does not refer to any specific NP as it does in its anaphoric usage, but to a larger portion of the text.

Blackfoot frequently uses phrases containing discourse deictic demonstratives as formulaic endings to stories as in 11. This example is the conclusion to a short story about some young men who play a trick on a group of old women.

(11) **Kɛ́nniaie**   nanístsksinoàii   ámoksisk   kipitákeks.
     ki-ann-yi-ayi   n-aanist-ssksino-aa-yi   am-o-iksi-hka   kipita-akíí-iksi
     and-DM-IN.SG-ID   1-MNR-know.TA-DIR-3PL   DP-INT-3PL-INVS   aged-women-3PL
          '**And that is** all I know about these old women.'   Uhlenbeck (1912:204)

In this example, the discourse deictic *kénniaie* refers to the entire preceding story, linking it with the proposition that the speaker has shared all his knowledge on the topic.

In the textual data I analyzed, Blackfoot discourse deixis is always anaphoric and never cataphoric. While further investigation may uncover instances of cataphoric uses, thus far there

is no evidence from Blackfoot texts to support the claim that cataphoric reference in discourse deixis is universal.

3.4. RECOGNITIONAL USES**.** Recognitional uses serve to activate information that is new to the discourse, but already known to the addressee. Frantz (2009:64) describes the *ann* stem as encoding 'familiarity to the addressee' in addition to proximity features. I propose that notions of familiarity arise from the use of the *ann* stem in recognitional uses. Thus, the distance features conveyed by *ann* in situational uses and the familiarity to the speaker expressed by *ann* in recognitional uses are not simultaneously expressed in each use of the stem *ann* but are a result of two different pragmatic functions.

In the textual data I examined, *annáhka* is the only form that appears in recognitional uses. In other words, recognitional forms are instances of the *ann* stem with the RRC suffix *-hka* and without the geometric configuration suffix *-o*. Further research is needed to determine whether recognitional demonstratives only occur with proximate animate nouns, or whether this is a result of a limited corpus and the relative rarity of this pragmatic use.[5]

Recognitional uses are sometimes difficult to identify because it is not known what information is common knowledge to the narrators of Uhlenbeck's (1912) stories and their audiences. The portions of the texts that do contain identifiable recognitional uses often involve two characters within the narrative engaged in dialogue. The recognitional uses described below are all from such contexts in which (i) a demonstrative form is used to introduce new information in an exchange between two characters within the narrative and (ii) the narrator has indicated that the referent of the demonstrative is already known to the addressee prior to its mention. One such example is given below in 12. In this example, a woman has indicated that she loves a man named Round-cut-scabby-robe. Her public confession of affection causes Round-cut-scabby-robe embarrassment and so he leaves. As he does so, he asks a friend to meet him later and relay what the women says about him after he leaves. When Round-cut-scabby-robe reunites with his friend, he asks the following question.

(12) Áuke, tsánìu   **annáχk**   àkéuaχk.
     óki   tsá   waanii-wa **ann-wa-hka**   akíí-wa-hka
     DISC   what say.AI-3SG **DM-3SG-INVS** woman-3SG-INVS
     'Now, what did **that** woman say?'   Uhlenbeck (1912:73)

The use of *annáχk* to refer to the woman whom they had previously discussed is not characteristic of tracking usage since she is not mentioned in the discourse that begins when the two friends reunite. The referent (the woman) is known to both discourse participants, but is new, within the discourse event.

Another example of recognitional use comes from a story in which a girl befriends a bear in the forest. In this story, the girl's 'meddlesome' younger sister tells their father that her older sister is always playing with a bear. The father gathers a hunting party and kills the bear. The older sister blames her younger sister for the bear's death and sends her on an errand that will result in the younger sister's demise. Part of her instructions is given in 13.

(13) Istɑ́pot   **omím**   itsínitɑχpi   **annáχk**   kyáio
      itap-oo-t   **om-yi-ma**   it-i'nit-hp-yi   **ann-wa-hka**   kiááyo-wa
      toward-go.AI-IMPV   **DD-IN.SG-STAT**   DCT-kill-CN-IN.SG   **DM-3SG-INVS**   bear-3SG
        'Go **over there** where **that** bear was killed.'   Uhlenbeck (1912:103)

Here the referent (the dead bear) is known to both speech-act participants and the situational use of *omím* excludes the possibility of situational interpretations of *annáχk*, since they employ two different stems to refer to the same location. The bear had not been previously mentioned in any conversation between the two girls, so it is new to their discourse. Since the last overt mention of the bear in the narrative is ten clauses prior to 13, even within the context of the overall narrative, this is not a tracking use of *annáχk* as *om* is used for long-distance tracking (§3.2).

4. CONCLUDING REMARKS AND FURTHER RESEARCH. Himmelmann's (1996) four pragmatic uses of demonstratives are each attested in Blackfoot. This lends support to Himmelmann's first claim that all four pragmatic functions are universal. However the Blackfoot data do not support Himmelmann's second claim that all four uses are equally basic. While the Blackfoot demonstrative system employs three stems in situational uses, only the *om* and *ann* stems are used non-situationally. The *ann* stem is the only one used in all four pragmatic functions. The restricted distribution of the *am* and *om* stems in non-situational functions supports Diessel's (1999) argument that situational uses are the basic use from which the others are derived. Furthermore, the obligatory presence of the RRC suffix *-hka* on recognitional demonstratives indicates that recognitional forms are morphologically more complex than situational forms. Diessel, citing markedness theory (Croft 1990), states that morphologically more complex forms are derived from morphologically less complex forms. Thus textual data from Blackfoot supports two of Diessel's arguments for the special status of situational uses as basic and the derived status of non-situational uses.

      As noted above, the attested discourse deixis occurrences in the textual data are all anaphoric. It is unclear whether this is because cataphoric discourse deixis usage does not occur in Blackfoot, or whether it is an accident of the data as a result of a relatively small corpus. Further research is needed in this area, but not only on Blackfoot; discourse deixis usage is undescribed in many languages, even those consider well-documented.

      The relationship between topic/focus and demonstrative usage is another area that needs further investigation. It may be that both *om* and *ann* are part of the focalization system, or that *ann* encodes topic and *om* focus. To address these questions, it is necessary to investigate instances of focalization and topicalization both with and without demonstratives to determine the contexts in which each strategy is used.

REFERENCES

BLISS, HEATHER. 2005. Topic, focus, and point of view in Blackfoot. Paper presented at WCCFL 24, Simon Fraser University, Vancouver, Canada.

CLEARY-CAMP, JESSICA. 2007. Universal uses of demonstratives: evidence from four Malayo-Polynesian languages. *Oceanic Linguistics* 46.325–347.

COMRIE, BERNARD. 1989. *Language universals and linguistic typology.* Chicago: University of Chicago Press.

CROFT, WILLIAM. 1990. *Typology and universals*. Cambridge: Cambridge University Press.

DIESSEL, HOLGER. 1999. *Demonstratives: Form, function and grammaticalization*. Amsterdam: John Benjamins.

FILLMORE, CHARLES. 1997. *Lectures on deixis.* Stanford: CSLI Publications.

FRANTZ, DONALD G. 2009. *Blackfoot grammar* (2nd edition). Toronto, Ontario: University of Toronto Press.

GEERS, GERARDUS JOHANNES. 1917. *The adverbial and prepositional prefixes in Blackfoot.* Leiden: N. V. Boekdrukkerij.

GENEE, INGE. 2005. Blackfoot demonstratives. Invited presentation at The University of Alberta Linguistics Colloquium, Edmonton, Alberta. 1 April 2005.

GLOUGIE, JENNIFER R. S. 2000. *Topics in the syntax and semantics of Blackfoot quantifiers and nominals.* Vancouver, BC: The University of British Colombia MA Thesis.

HIMMELMANN, NIKOLAUS P. 1996. Demonstratives in narrative discourse: A taxonomy of universal uses. *Studies in anaphora*, ed. by Barbara Fox, 205–54. Amsterdam: John Benjamins.

IMAI, SHINGO. 2003. *Spatial deixis*. Buffalo, NY: SUNY dissertation.

LAKOFF, ROBIN. 1974. Remarks on 'this' and 'that'. *Chicago Linguistics Society* 10.345–356.

LAMBRECHT, KNUD. 1994. *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents.* Cambridge: Cambridge University Press.

LYONS, JOHN. 1977. *Semantics,* (vols. 1 & 2). Cambridge: Cambridge University Press.

SCHUPBACH, S. SCOTT. 2013. Situational demonstratives in Blackfoot. *Coyote Papers* 21. Online: http://arizona.openrepository.com/arizona/handle/10150/270993.

TAYLOR, ALLEN. 1969. *A grammar of Blackfoot*. Berkeley, CA: UC Berkeley dissertation.

UHLENBECK, CHRISTIANUS C. 1912. A new series of Blackfoot texts: from the Southern Peigans Blackfoot Reservation, Teton County, Montana. *Verhandelingen der Koninklijke Akademie van Wetenschappen te Amsterdam, Afdeeling Letterkunde, Nieuwe Reeks* 13.1. Amsterdam: Johannes Müller.

UHLENBECK, CHRISTIANUS C. 1938. A concise Blackfoot grammar. Amsterdam: North Holland Publishing Company.

VAN VALIN, ROBERT D., and RANDY J. LAPOLLA. 1997. *Syntax: structure, meaning, and function*. Cambridge: Cambridge University Press.

[1] 3=animate proximate (3rd person); AI=animate intransitive verb stem; CN=conjunctive nominal; DCT=deictic preverb; DD=distal demonstrative stem; DIR=direct; DISC=discourse marker; DM=medial demonstrative stem; DP=proximal demonstrative stem; DTP=distinct third person pronominal suffix; DUR=durative marker; ID=identificational suffix; II=inanimate intransitive verb stem; IN=inanimate noun; INT=interior geometric configuration; IMPV=imperative; INVS=invisible; LOC=localizer; MA=motion away; MNR=manner preverb; MT=motion toward; NOM=nominalizer; PL=plural; SG=singular; STAT=stationary; TA=transitive animate verb stem.

[2] Geers (1917:100) lists *ot-* and *aut-* as 'cognate with' *oto-* 'go to do (something)'.

[3] This analysis is based on the definition of focus given by Lambrecht (1994:207) that focus is the element in a proposition's assertion that differs from its presupposition. Comrie (1989:63) describes 'focus' as 'the essential piece of new information carried by a sentence' and topic as 'what the sentence is about'. Bliss (2005) argues that Blackfoot uses proximate/obviative morphology to mark focus and topic. Since the *om* stem takes both proximate and obviative morphology, it is likely that tracking uses of *om* serve some other function than to mark topic or focus.

[4] Further quantitative research focused specifically on comparing the recurrence rates of new information introduced by *om* with new information introduced without *om* is needed to confirm this initial observation.

[5] Since, by definition, recognitional uses introduce information that is new to the discourse, they frequently function as the focus of their clause. Based on Bliss's (2005) proposal that proximate forms are used to encode focus, it is possible that recognitional forms are always proximate.

Determining prefab status in New Mexican Spanish, the case of *decir*

Víctor A. Valdivia-Ruiz
University of New Mexico

ABSTRACT. The present study addresses the issue of how to determine what a prefab is; specifically, it proposes features such as frequency, morphosyntactic variability, and prosodic distribution, constitute operational measures with which structural gradience can be accounted for. Results from the analysis of *decir* clauses from the *New Mexico and Colorado Spanish Survey* (Bills & Vigil 2008) show subject expression and position, verb's position, and prosodic distribution between the verb and its complement are important indicators that a particular subject-tense collocation is a prefab.

1. INTRODUCTION. Prefabs are multiword sequences, conventionalized expressions 'known to speakers as such' (Bybee and Beckner 2009:835), but whose meaning may or may not be idiosyncratic. These complex units are of particular interest within functional-cognitive approaches to linguistics, especially within the USAGE-BASED THEORY, not only because they make evident the need to develop a theory that deals with pairings of form-meaning larger than the word, but also because of their relevance in the general-domain of analogy. Since speakers' experiences affect both categorization of linguistic input and production of new utterances (Bybee 2010:8), prefabs provide a large repertoire of elements for comparison given the vast and complex networks they can be part of by establishing links with other instances both through their conventionalized meaning as a whole and through their internal lexical components.

As with all other constructions, prefabs result from chunking, a process by which the repetition of two or more words together causes them to be interpreted and later produced, as a whole unit. The central role of repetition implies the process is gradual, so linguistic units move along a continuum from one category to another, and it is not always easy to distinguish the boundaries of many categories of grammar. In the case of prefabs, gradience allows us to include in the category semantically transparent expressions such as 'black and white' but also others as 'to rub the wrong way'*,* which cannot be understood by simply adding the individual meanings of each one of its parts. Because prefabs are gradually developing units, the emergence of a new one does not mean the old one disappears; they coexist as long as they both continue being used in discourse, and the boundaries between both of them are blurred. Furthermore, it is often the case that some of the lexical meanings of one or more items in the prefab are still observable, which allows them to establish links with other constructions containing semantically similar items (Bybee & Torres 2009).

Furthermore, constructions may 'include not only properties of the situation described by the utterance but also properties of the discourse in which the utterance is found […] and of the pragmatic situation of the interlocutors' (Croft 2001:14). Thus, the more a prefab is used in particular communicative situations, the more semantic bleaching it experiences, and, consequently, the more likely it will convey pragmatic and even epistemic nuances.

Based on the former characteristics and phenomena, I will argue there is evidence suggesting some multi-word sequences involving *decir* are developing prefab status, and that the former can be determined by a quantitative analysis of various morphosyntactic and prosodic features such as subject expression and position; complement type, realization and position; co-

occurrence of the verb and its arguments in the same intonation unit; and complement's distribution in one or more intonation unit, among others. Criteria for the operationalization of these features will be discussed in detail in section 4.

2. DATA. Data for this paper were extracted from 18 one-on-one interviews from Bills & Vigil's 2008 *New Mexico and Colorado Spanish Survey* (NMCOSS). The duration of speakers' formal instruction varies from minimal schooling to 17 years, and their professions are very diverse. Data from interviewers have also been included in the study because, as members of the community to be studied, they provide valuable data as their speech may show discourse functions only observable in their interaction with the interviewees' speech; for instance, the common use of questions and their goal of eliciting speech from the interviewee. It is important to mention that code-switching is not unusual in the interviews.

　　Interviews were transcribed according to Du Bois and colleagues' 1993 system, in which each line represents an intonation unit (IU), a speech segment uttered under one intonation contour.[1] Even though IU's do not directly correspond to syntactic units such as phrases or clauses, they do appear to play a role in the way grammar is organized (Chafe 1994, Ono & Thompson 1995) and thus are incorporated into the analysis. Using the software MonoConc (Barlow 1999), data were extracted from approximately 8 hours, 10 minutes of recording, corresponding to 87,800 words and 24,544 IU's.

3. METHODOLOGY. The sample was extracted following two criteria. The first one was to exclude clauses in which the verb or its arguments were unclear, or incomplete.[2] Both phenomena are illustrated by 1 below:

(1)　　Cuando se fueron,
　　　　... (2.0) dijo,
　　　　<X hallo el X>--
　　　　　　'When they left, she said, <X I find the X> --'[3]

　　The second criterion was to have a similar amount of tokens from each interview. In order to reach 300 tokens, approximately 17 from each of the 18 interviews were needed. However, some interviews did not have 17 tokens, and thus, more than 17 tokens were taken from others. The minimum extracted from one interview was six, and the maximum 19 tokens. For interviews with more than 17 tokens available, the first 17 clauses were extracted.

　　Each clause was coded for a series morphosyntactic and prosodic features (see section 4) in order to identify both the most frequent collocations and their structural variability. The latter is particularly important in cases in which two or more collocations are constructed around the same verbal form since it provide clues regarding whether they all represent the same prefab, or whether they are indeed different ones.

4. CODING. For the analysis, the basic units are the clause and clause combinations, understood as the syntactic structure formed by a verb and its arguments. This follows previous research on structural patterns (e.g. Scheibman 2001, 2002; Travis 2006), which will allow a clearer and more precise comparison between those results and the ones from this study. In order to identify what factors affect morphosyntactic variability, as well as the role of cognitive management of information, each clause was coded for the following features:

a. Subject person and number (1SG, 2SG, 3SG, 1PL, etc.)[4]
b. Subject realization (pronouns, full nominal phrases, etc.)
c. Subject position (SV or VS)
d. Verbal structure (simple, modal, and periphrastic clauses)
e. Tense, aspect, mood (TAM)
f. Type of complement[5]
g. Complement's position (VC, CV, CVC)
h. Verb and complement's prosodic distribution[6]
i. Complement's prosodic distribution[7]
j. Indirect object (occurrence vs. non-occurrence, morphological realization)

5. RESULTS. Three verbal tenses account for most of the clauses in the data: present (122/300, 40.6%), preterit (95/300, 31.66%) and imperfect (66/300, 22%); however, only the first group gave rise to collocations developing prefab status, even if their token frequency is not particularly high.[8] These collocations are: (*yo*) *digo* 'I say', *digo yo* 'I say, lit. Say-I', *como* (*LE*) *digo* 'as I say / as I told you / I'm telling you', and *le dicen* X 'they call it X'.

These results do not mean the other groups did not produce collocations with certain discourse functions, but they constitute more schematic (i.e. general) patterns. For instance, although (*él/ella*) *dijo* 'he/she said' is the most frequent item in the data (54/300: 18%), as well as the central exemplar for both preterit (54/95: 56.84%), and 3SG clauses (54/92: 58.69%), it shows a high degree of morphosyntactic variability; hence, there is not a specific collocation strong enough to cause structural fixation.[9]

5.1. (*YO*) *DIGO* 'I SAY'. This item represents 53% (26/49) of the 1SG-present group. It includes both unexpressed (21/49: 43%) and expressed subjects (5/49: 10%) in preverbal position. The preponderance of *(yo) digo* is observed too in the oral section of the database *Corpus del Español* (Davies 2002), where it is the second most frequent form for *decir* in present tense*: digo* appears 2,432 times, and *yo digo* 422 times. Together, these forms represent 32.5% (2854/8784) of the group.

In most cases, the collocation has a direct object (22/26: 85%), usually realized by a clause, or a sentence (19/26: 73%). These finite complements tend to appear (14/19: 74%) in the canonical postverbal position. Regarding indirect objects, they appear when the speaker uses *(yo) digo* 'I say' to refer to an utterance from a previous speech event in which there was a specific addressee:

(2)   Yo **les** digo,
        Yo no nací pa' sentada. (219-1A2: 224)
               'I tell **them**, I wasn't born [to be] sitting.'

When the indirect object is not included, the utterance is not addressed to anyone in particular, not even the interviewer, because the speaker is expressing a belief, opinion, or decision, as in 3 and 4; thus, *decir* 'to say' does not reproduce a speech event, but a thought. Given the variation of meaning associated with the presence or absence of an indirect object pronoun, it is plausible to think two different prefabs are emerging from the collocation (*yo*) *digo* 'I say'. More data are necessary to confirm the hypothesis.

(3)   yo también digo que **gavilán es hawk**. (270-1B2: 93)
         'I too say that ***gavilán* means hawk**'

(4)   B: What advice would you [leave them]?
         W: [Pues yo],
             yo digo=,
             .. que yo,
             todo lo que tengo,
             yo --
             yo me ha tenido --
             yo ha tenido que trabajar. (190-3B2: 19)
                 B: 'What advice would you leave them?'
                 W: 'Well, I, I say that I, **everything I've, I, I've had, I've had to work** [for it].'

A feature that in this data is shown to be central to determining whether a collocation is developing prefab status is the prosodic distribution between the verb and its complement. More than half of the complements for (*yo*) *digo* 'I say' (15/26: 58%) appear in a different intonation unit than the verb, and most of them are finite complements (14/19: 74%), as in 5. If (*yo*) *digo* 'I say', being a low content verb, does not convey new information, then what is the cause for such prosodic independence?

(5)   Yo les digo,
         Yo no nací pa' sentada. (219-1A2:224)
                 'I tell them, **I wasn't born [to be] sitting.**'

For several of the tokens, the explanation lies in the use of the prefab as a direct speech device. This device attempts to eliminate the cognitive effects associated with the displaced mode in the discourse, the mode in which the speaker focuses on experiences that originated from past events (Chafe 1994:199). Moving from this mode to the immediate mode means the speaker pretends that the experience is perceived either as ocurring at the time of its representation, or as relevant to it. The effect is stronger when direct speech is used together with historical present, as in 6.
         Furthermore, the immediate move is where the speaker can access, evaluate, and regulate the idea being communicated, instead of merely remembering or imagining, as in 5 above. In the interview, the speaker is talking about how her kids want her to stop weaving and farming because of her age. Thus, when she says *yo no nací pa' sentada* 'I wasn't born to be sitting', she is offering an explanation for her refusal to follow her kids' wishes. Additionally, the utterance does not refer to a single previous speech event, but rather to an idea expressed habitually. The important thing here is not whether the speaker says exactly the same thing time after time, but the repetition of an idea. Thus, when using direct speech, the speaker is performing two processes at the same time, one with the verb and the other with the complement. Since each of them has its own cognitive load, they tend not to share an intonation unit.
         In the data, 63% (12/19) of all (*yo*) *digo* 'I say' clauses with finite complements show this prosodic structure, suggesting such subject-verb collocations often perform one of the functions described above. Moreover, since the event is presented as if occurring at the time of its expression, it exists by itself, therefore it is not syntactically dependent on *(yo) digo* 'I say'. Such

a loss of dependence allows cases in which the verb appears inside the complement, or in which two instances of *digo* 'I say' refer to the same complement, as in 6 and 7, respectively.

(6)   quítate el cute y,
        .. <u>No</u>,
        le digo,
        <u>ya me voy pa' mi casa</u>. (318-1A3: 160)
                '[He said] Take off your coat. <u>No</u>, **I tell him**, <u>I'm going home</u>.'

(7)   (H) Y luego cuando acabó con mi hermana **le digo**,
        <u>pues</u>,
        eh,
        <u>no me puede pegar a mí porque yo no estoy con usted</u>, **le digo**. (318-1A3: 143)
                'And then, when he was done beating my sister, **I tell him**: "well, you have no right to hit me because I do not live with you", **I tell him**.'

Although few in number, utterances such as 6 and 7 are very relevant for the argument about the prefab status of (*yo*) *digo* for two reasons. Firstly, the fact that the collocation is intercalated into the complement indicates loss of syntactic cohesion between both constituents; that is to say, they seem to be structurally independent from each other rather than two elements of a sentence. Secondly, structural changes have semantic and functional effects. Thus, using (*yo*) *digo* 'I say' when the information was already or is being presented implies the collocation is doing more than merely reporting information. For instance, in 6 and above, by recreating the actual wording of the original discussion, the speaker also indicates his involvement and stance (cf. Chafe 1994). Moreover, in intra-clausal position, the collocation can convey epistemicity without having a textual function, as in 8, where *yo digo* has a meaning closer to 'I mean' than to 'I say.'

(8)   A: Yo también lo <X digo para decir X>,
            (H) porque nos pasamos tanto tiempo en el teléfono,
            digo,
        Y: Sí.
        A: <u>Y luego no caminamos</u>, (147-2A2: 133)
                A: 'Me too, <X I said just as an excuse X>, because we spent a lot of time on the
                    telephone, **I mean**,'
                Y: 'Yes'.
                A: '<u>and then we don't go for our walk</u> …'

In a previous part of the conversation, speaker A is talking about one of her friends who recently bought a cordless telephone. Then, she states she wants one too (first line) and offers the reason for the potential purchase (second line) followed by *digo* 'I mean' (lit. 'I say'). Finally, after an interruption by the other person, Alicia mentions the consequences of spending too much time on the telephone (fifth line). The complexity of this segment of conversation clearly illustrates that of *(yo) digo* 'I say'. First, it is not clear whether the collocation refers to the clause in line 2 or to the one in line 5, or whether both clauses compose a whole explanation. Second, although the situation expressed by Alicia (i.e. spending too much time on the telephone) is

147

something habitual and relevant to the present, it really does not refer to a specific speech event, so *(yo) digo* does not function as direct or indirect speech. Third, even if we consider both lines 2 and 5 as parts of one explanation, the intercalation of *(yo) digo* suggests the speaker uses it to indicate that there is relationship between both clauses. In this sense, the collocation has a function similar to English 'I mean'.

Finally, in all cases of intra-clausal position, the verb appears with an unexpressed subject and without an indirect object. Consequently, it could be possible that *digo* 'I say' is emerging as a new construction from the prefab *(yo) digo* ALGO 'I say something'. Unfortunately, there are not enough tokens in the data to test this hypothesis.

In sum, evidence supporting the existence of the prefab *(yo) digo* 'I say' comes from its token frequency, its predominant use to introduce quotes, its tendency to appear in its own intonation unit, and the possibility to occur in different positions within the clause.

5.2. *DIGO YO* 'I SAY' (LIT. 'SAY I').[10] *Digo yo* represents 10% (5/49) of all 1SG-present clauses, a proportion equal to that of *yo digo* 'I say' (5/49: 10%). Certainly, frequency does not provide enough evidence to support the status of *digo yo* as a construction different and independent from the prefab *(yo) digo* 'I say'; however, it does suggest a tendency for the subject to be expressed after the verb. Moreover, considering VS is not the typical word order for transitive clauses, the subject's postverbal position relates to a particular discourse function.

Digo yo 'Say-I' mostly introduces into the discourse utterances that the speaker has expressed in previous events; nevertheless, complements occur in both preverbal and postverbal position, as shown in 9 and 10, respectively. The only case in which is not clear whether *digo yo* 'Say I' has a complement or not is in 11. On the one hand, the clause seems to be truncated by *y yo sé que te vas adelantando* 'and I know you get ahead'; on the other hand, it looks as if the former were indeed an explanation introduced by *digo yo* 'Say I'.

(9)   Ellos sabían que me –
        .. tenían que respetarme.
        .. y asina he sido siempre,
        le **digo yo**. (311-3B2: 311)
                'They know they.. they had to respect me, and I've been like that, **I am telling you.**'

(10)   (H) Y por eso **digo yo** que,
        If you do that,
        .. y el mismo patrón de él,
        (H) .. he realizes that. (190-3B2: 177)
                'And that's why **I say** that, if you do that, and his boss himself, he realizes that.'

(11)   [y por eso],
        eh **digo yo** y,
        y yo sé que .. te vas adelantando,
        % % te ayuda mucho,
        en otras palabras. (190-3B2: 105)
                'and that's why, eh **I say**, and I know you get ahead, in other words, it helps you a lot.'

Prosodically, *digo yo* and its complement always appear in different intonation units, as in the examples above. Such a distribution could be related to the fact that speakers always use the collocation to say something about a situation, not to present a speech event per se. That is to say, the collocation does not report actual speech, but it presents an utterance summarizing, or explaining the event being talked about. In 9, for instance, *y asina he sido siempre* is not exactly a quote from the dialogue between the interviewer and the interviewee, but some kind of justification by the latter about his behavior towards his grandchildren.

5.3. *COMO* (*LE*) *DIGO* 'AS I SAY/TELL (YOU)'. Although this collocation appears only seven times in the data, its morphosyntactic features suggest its prefab status. It represents 14% (7/49) of all 1SG-present tense occurrences of *decir*, 7.5% (7/93) of all 1SG occurrences, and 6% (7/122) of all present tense occurrences. In most of the tokens, the collocation includes an indirect object pronoun referring to whom the linguistic utterance is addressed.

Additionally, *como* (*LE*) *digo* 'As I say/tell (you)' always appears with unexpressed subjects and it takes clausal complements. However, such complements are not typical because they really do not form a syntactic unit with the verb. Rather, the collocation is intercalated into the information being presented as in 12. Even when *como* (*LE*) *digo* 'As I say/tell (you)' precedes a whole sentence, it seems to be used as a discourse connector between a sentence and the previous one, as in 13.

(12)   Y yo soy their older sister.
        so yo los --
        yo los cuidaba,
        yo los cuidaba y,
        todavía cuando me casé,
        como digo,
        y estaba --
        me enfermaban mis babies y yo,
        y aquí estaba cuidando, (117-1A3)
                'and I am their older sister. So I-- I took care of them. I took care of them and,
                even after I got married, as I was saying, and she was -- .. if my babies got sick, I
                was here taking care of them.'

(13)   H: no sabía cuándo parar de jugar.
        F: [Hm]
        H: **[Como] te digo** <u>ese hombre</u>,
            .. ganó.
            Ganaba .. doscientos mil,
            trescientos mil.
            .. Pero se levantaba de la mesa,
            se levantaba pura madre. (156-1A1: 279)
                H: 'He didn't know when to stop gambling.'
                F: 'Hm.'
                H: '**I'm telling you**, that man won -- he would win two hundred thousand, three
                    hundred thousand, and not even then, he would walk away from the table.'

The relation between *como* (*LE*) *digo* 'As I say/tell (you)' and the information it refers to seems to be related to the prosodic distribution of both constituents. Except for 13, no elements from the complement share the intonation unit with the collocation. This fact is relevant because the speaker presents the information as if it were already known to the interlocutor, yet it does not appear in the verb's intonation unit. Thus, prosody suggests the collocation is performing a specific function in the discourse, namely to reactivate a piece of information or to present new information as if it had been previously introduced into the discourse. This process allows the speaker to use a statement as a part of a whole argument or explanation for a particular event. For instance, in 14, the speaker declares she could not speak English fluently as a fact known to her interlocutor, but her intention is to make the point that such circumstances did not impede her and her family from moving forward in life.

(14)   O: Y=,
        como digo,
        no lo aprendería a hablarlo tan fluently,
        .. you know? (117-1A2: 127)
    A: Mhm.
    O: inglés.
        But,
        nunca nunca nos atrasamos.
            O: 'And, **as I say**, I might not have learned to speak fluently, you know?'
            A: 'Mhm.'
            O: 'English. But we never got behind in life.'

5.4. *LE*(*S*) *DICEN* X 'THEY CALL IT/THEM X'. One of the clearest prefabs with *decir* 'to say' is *le*(*s*) *dicen* X 'they call it/them X', illustrated in 15. The collocation represents more than half (15/28: 53.5%) of the 3PL$_{gen}$-present tense occurrences of *decir*, as well as a significant proportion of both 3PL$_{gen}$ subjects (15/58: 26%) and present tense clauses (15/122: 12%).

(15)   [Rastros] **les dicen**,
        o **les dicen** huellas.
        [no sé]. (214-1B2: 143)
                'Traces, **they call them**, or **they call them** footprints. I don't know.'

As observed in the above example, the collocation has a generic subject, mostly unexpressed (13/15: 86.6%), and an indirect object pronoun referring to an object or place whose name is being asked to the speaker. The name is typically expressed by a bare noun in postverbal position (10/15: 66.6%), as *huellas* 'footprints' in 15 above.

On the topic of prosody, the indirect pronoun, the X complement, and the verb share the intonation unit in practically all of the tokens (13/15: 86.6%). Since the complement is the new information introduced by the speaker, such prosodic distribution is only possible if *dicen* 'they say' has a low cognitive load. However, the prefab seems to perform a specific discourse function in the discourse; speakers use it to indicate how people from other communities call a particular object, animal, or place. The prefab, then, is a kind of evidential marker, allowing the speaker to present information without being responsible for its accuracy as shown in 16:

(16)   [No sé cómo se] llamarán.
        Te- --
        .. <u>Tepocates</u> les **dicen** en español. (214-1A2: 37)
                'I don't know what their name is. Te-- <u>tepocates</u>, **they call them** in Spanish.'

This prefab is a clear example of how new constructions emerge from previously established ones. In this case, *le*(*s*) *dicen* X 'they call it/them X' serves as a model for *le*(*s*) *decían* X 'they called it/them X', as in 17. This prefab appears eight times in the corpus, which represents just 2.6% of the whole data; however, it accounts for 14% (8/58) of all 3PL$_{gen}$ subjects, and 27.5% of all 3PL$_{gen}$-imperfect clauses (8/29)

(17)   Y hay otras blancas que se dan <X ahí X>.
        les dicían <u>tanapes</u>.
        .. unas [blancas]. (10-3A2)
                'and there are some white ones that grow there, **they called them** <u>tanapes</u>
                [turnips]. Some white ones.'

6. CONCLUSIONS. The analysis of the data demonstrates that different prefabs can emerge not only from the same verb, but also from the same collocations. In additon, it shows that not only morphosyntactic features, but also prosodic ones provide a set of operational measures to determine whether a particular collocation constitutes a prefab or, rather, a general structural pattern.

In addition, the analysis provided evidence regarding the role of established prefabs in the development of new ones. Thus, the loss of morphosyntactic variability of *digo* 'I say' allows the prefab to move to sentence-final, and even to an internal position, and to develop new semantic and pragmatic nuances. The phenomenon is also observed in the emergence of *le*(*s*) *decían* X 'they called it/them X', which is related syntactically, semantically, and pragmatically to *le*(*s*) *dicen* X 'they call it/them X'.

REFERENCES

BARLOW, MICHAEL. 1999. *Monoconc*. Houston, TX: Athelstan.

BILLS, GARLAND and NEDDY VIGIL. 2008. *The Spanish of New Mexico and Southern Colorado: A linguistic atlas.* Albuquerque, NM: University of New Mexico Press.

BYBEE, JOAN, 2010. *Language, usage and cognition.* Cambridge: Cambridge University Press.

BYBEE, JOAN, and CLAY BECKNER. 2009. Usage-based theory. *The Oxford Handbook of Linguistics analysis,* ed. by Bernd Heine and Heiko Narrog, 827–855. Oxford: Oxford University Press.

BYBEE, JOAN, and RENA TORRES CACOULLOS. 2009. The role of prefabs in grammaticalization: How the particular and the general interact in language change. *Formulaic Language: distribution and historical change* (vol. 1)*,* ed. by Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali, and Kathleen M. Wheatley, 187–217. Amsterdam: John Benjamins.

CHAFE, WALLY. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing.* Chicago: The University of Chicago Press.

CROFT, WILLIAM. 2001. *Radical construction grammar: syntactic theory in typological perspective.* Oxford: Oxford University Press.

DAVIES, M. 2002. *Corpus del Español: 100 million words*, *1200s–1900s*. Online: http://www.corpusdelespanol.org.

DU BOIS, JOHN; STEPHAN SCHUETZE-COBURN; SUSANNA CUMMING; and DANAE PAOLINO. 1993. Outline of discourse transcription. *Talking data: Transcription and coding in discourse*, ed. by A. Edwards and Martin D. Lampert, 45–89. Hillsdale, NJ: Lawrence Erlbaum.

ONO, TSUYOSHI, and SANDRA THOMPSON. 1995. What can conversation tell us about syntax? *Alternative linguistics: Descriptive and theoretical modes,* ed. by Philip W. Davis, 213–271. Amsterdam: John Benjamins.

SCHEIBMAN, JOANNE. 2001. Local patterns of subjectivity in person and verb type in American English conversation. *Frequency and the emergence of linguistic structure*, ed. by Joan Bybee and Paul Hopper, 61–89. Amsterdam / Philadelphia: John Benjamins.

SCHEIBMAN, JOANNE. 2002. *Point of view and grammar. Structural patterns of subjectivity in American English conversation.* Amsterdam: John Benjamins.

TRAVIS, CATHERINE E. 2006. Subjetivización de construcciones: Los verbos 'cognitivos' en el español conversacional. *Memoria del VIII Encuentro Internacional de Lingüística en el Noroeste* (vol. 2)*,* ed. by Rosa Ma. Ciscomani Ortiz, 85–109. Hermosillo: UniSon.

NOTES

[1] A period after an IU indicates final transitional continuity; a colon indicates continuing transitional continuity, and a question mark indicates an appeal (cf. Du Bois et al: 1983).

[2] In Du Bois et al (1983), <X X> indicates unclear speech, and -- indicates a truncated IU. Other features indicated in this paper are truncated word '-', brief pauses '..', medium pauses '…', long pauses '…(N)', speech overlap '[ ]', and audible exhalation '(H)'.

[3] For the examples' translation, punctuation is used in its regular way.

[4] Generic subjects are indicated with 'gen'.

[5] It codes whether the clause has a complement and, if so, the type of such constituent. I am using the term 'complement' broadly, to refer to any morphosyntactic structure occurring with the verb to express a whole idea, without focusing on its syntactic function. Thus, coded under this category are not only direct object complements, but also predicative ones, as in examples (3) and (15) respectively.

[6] Whether the verb and its complement occur in the same intonation unit. Clauses are marked as such when the verb and at least a lexical element of the complement share the same IU.

[7] Whether the complement is expressed in one or more IU's.

[8] In fact, token frequency is not an indispensable condition for prefabs, as stated by Bybee (2007:16): 'Prefabs are not especially high in frequency, and yet it is clearly their repetition that has given them their social status of conventionalized and their cognitive status of easily accessed routines.'

[9] Nevertheless, it is important to mention that, in the discourse, SBJ-*dijo* emerges as a quotative device; it presents speech from previous conversations. More importantly, in such a function, the collocation often appears in its own intonation unit, and it occurs not only before its complement, but also after it, and even shares the complement with another item of *decir*

[10] For this collocation, I will use 'Say I' to distinguish it from both (*yo*) *digo* and *yo digo* 'I say'. In the translations, I will use I SAY.

Repairing the common ground in a joint ASL narrative:
A mental space approach

Deborah Wager
University of New Mexico

ABSTRACT. Traditional explanations of conversational repair point to trouble spots in the words of a piece of discourse that speakers need to fix. In this paper mental spaces representations of the common ground in a signing dyad and between the signers and the audience are used to examine repairs in an ASL joint narrative. Repair sequences occur when the signers recognize a discrepancy in their mental space representation of the common ground between themselves or between the signers and the audience.

1. INTRODUCTION

1.1. REPAIR. Conversational repairs have been examined by many researchers over the last several decades, looking at such things as the structure of repair (e.g. Schegloff, Jefferson & Sacks 1977; Schegloff 2000; Svennevig 2008), ways it is used by children (e.g. Konefal & Fokes 1984, Salonen & Laakso 2009), and pragmatic uses of repair (e.g. Liebscher & Dailey-O'Cain 2003). However, the question of motivation behind a repair has been largely unaddressed, with researchers simply referring to 'trouble spots' in the discourse.

The traditional analysis of repair is that when either a speaker or a listener detects trouble in speech, they move to repair that trouble, usually in the same or the next turn (Schegloff, Jefferson & Sacks 1977). In most cases the speaker initiates the repair (self-initiated repair). The speaker can then complete the repair (self-completed repair), perhaps without the trouble ever being detected by a listener (e.g. revising word choice without altering the flow of signing), or the listener may complete it (other-completed repair). Other times a listener will bring attention to the trouble, usually on the next turn (other-initiated repair) and either repair it (other-completed repair) or allow the speaker to repair it (self-completed repair).

Clark (1996:153) describes four levels of contribution included in an utterance, any one of which can be implicated as a trouble source in the case of a repair.

- Level 1 is EXECUTION AND ATTENTION, making the sounds or motions involved in forming words, signs, and gestures, and seeing or hearing them. Trouble at this level would involve failure of the speaker or signer to articulate properly or, on the part of the interlocutor, to accurately see or hear.
- Level 2 is PRESENTATION AND IDENTIFICATION, saying or signing the utterance as an utterance, and recognizing the words or signs. Trouble at level 2 might involve mis-stating or not recognizing the utterance as an utterance.
- Level 3 is SIGNALING AND RECOGNITION, or MEANING AND UNDERSTANDING. Trouble at level 3 involves problems with understanding the meaning of the utterance.
- Level 4 is PROPOSAL AND CONSIDERATION of joint projects, which can be small (asking and answering a question) or large (moving to another city). Trouble at this level would involve the interlocutors having different goals.

155

Recognizing the source of trouble can be difficult, especially in self-repairs or when an utterance is aborted and a new utterance is presented in its place. This difficulty has been noted (Schegloff, Jefferson & Sacks 1977:363), saying that, 'In view of the point about repair being initiated with no apparent error, it appears that nothing is, in principle, excludable from the class "repairable".'

1.2. COMMON GROUND. Common ground is a concept that has sometimes been given different nomenclature such as COMMON KNOWLEDGE, MUTUAL KNOWLEDGE, or JOINT KNOWLEDGE (Clark 1996:93). It refers to the shared knowledge that exists between two people, which they understand to be shared. Common ground is built up between two people as they have shared experiences, which include both shared real-life experiences and shared descriptions of things that have happened when they were separate. If two people both know something but neither has reason to believe that the other knows it, it is not counted as common ground. Similarly, one can be mistaken as to what is in the common ground and believe that something is common ground when it is not. Sometimes these differences go undetected, and sometimes they become apparent during a piece of discourse.

In a conversation, each piece of information is presented as a potential addition to the common ground and must be accepted by the other participant before it is considered part of the common ground (Clark & Schaefer 1989). This ACCEPTANCE can be tacit, consisting of continued attention, or explicit, with the interlocutor verbally agreeing with the information, or somewhere in between.

A joint narrative has implications in looking at common ground because it necessitates reference to several different common grounds. When a single speaker produces a narrative the common ground is between the speaker and the listener. In a joint narrative we need to consider the common ground between the two speakers as well as between each speaker and the listener. If, for instance, two friends are telling another friend about an experience they had, one speaker might well turn to the listener and reference common ground between the two of them that is not shared with the other speaker.

1.3. MENTAL SPACES. According to Fauconnier's (1994) theory of mental spaces, speakers and listeners construct mental spaces that each contain their conceptualization of the reality for a different situation. Mental spaces can be constructed for times, places, domains (game, field of science, type of literature, etc.) (1994:31) or hypothetical situations.

An essential feature of mental spaces is that elements in a person's various mental spaces can represent the same person or thing and can be connected using the IDENTIFICATION (ID) PRINCIPLE, which states that if two elements of mental spaces are linked by a pragmatic function, a description of one may be used to identify the other (Fauconnier 1994:3). Pragmatic functions include such things as links between persons and names or persons and information about them. The ID Principle is important in that it links corresponding items in different mental spaces so that they can be discussed as single entities.

2. DATA. This study presents an analysis of the common ground construction and repairs in the joint narrative of a Deaf[1] couple introducing themselves and their family and telling of the birth of their first child (RSA Region V Project 2003). This video is part of a series of videos to train ASL-English interpreters for working in medical situations. The video of the Bartons[2] was separated into pieces, one for each individual introduction, one for the family introduction, and

one for the birth story of each of their five children. Despite this separation, there is contextual evidence in the segments to indicate that they were recorded in the order presented.

The joint project the Bartons undertake in this video is to provide information for interpreters so they will be better prepared to interpret for a childbirth. They complete this project by telling their story in great detail. The combined time of the eight video segments is 39:17, and the segment devoted to the story of the birth of their first child is the longest at 12:19. Both signers seem to be very willing to initiate repairs and to add information to the narrative, making this video appropriate for looking at the construction of mental spaces and their states at the time of repairs.

The video showed both signers simultaneously, making it possible to see when attention was being paid to the other signer. This is crucial for determining that each utterance was accepted into the common ground by the other signer because the extended turns of the narrative mean that explicit acceptance is rarely given. For long stretches of the narrative, continued attention is the only acceptance given.

The only common ground that is assumed to exist with the viewer at the beginning of the narrative is a knowledge of American Sign Language and general cultural background. The common ground between the signers and the viewer is built up as the narrative progresses. There is no separate common ground between the viewer and the individual signers, as the signers work together throughout the video to build up the common ground with the viewer.

The common ground between the signers is considerably more extensive, as they have been married for nine years and have five children, and they are telling the story of the birth of one of their children, an event they were both intimately involved in. However, some discrepancies arise involving the common ground between the signers, and these are the focus of repairs.

3. HOW COMMON GROUND IS CONCEPTUALIZED. Over the course of the joint narrative a variety of mental spaces are built and used by the signers. I will focus on five mental spaces for each signer: reality; a discourse space; an intention space; and two common ground spaces, one representing that person's common ground with their spouse and the other representing the couple's common ground with the viewers. The sequence and use of these spaces was developed by William Croft in a seminar entitled Mental Spaces and Discourse, taught in 2011 at the University of New Mexico (UNM).

3.1. REALITY SPACE. The most basic mental space is the REALITY mental space. For the signers, the Reality mental space includes the two of them, the set, and the camera, as well as the camera crew or any others who may have been in the room (see Figure 1). While each signer has a separate mental representation of reality, I will present a single schematic representation to limit the size and complexity of the figures. The camera (C) represents both the camera and, from the viewpoint of the signers, the generic viewer. For the viewer, Reality space involves the situation in which they are viewing the video (classroom, computer, projector, other students, or whatever may be present) as well as the contents of the video.
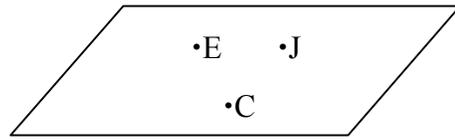
FIGURE 1. Reality space for the signers. Camera crew and any other people present at the filming are not included as they are not referenced and they may or may not be present.

3.2. DISCOURSE SPACE. The next mental space is the DISCOURSE SPACE. Each signer has a separate mental representation of what is being said (M_E and M_J) and I show these spaces separately, as in Figure 2. Each discourse space includes an INTENTION SPACE (I) so that not only what is being said but also the intended meaning is represented. Each of the initials repeated in the successive mental spaces should be connected by the ID Principle, as each represents the same person or entity in the successive spaces. For the sake of simplifying the diagrams these connections are omitted in the diagrams in this paper.
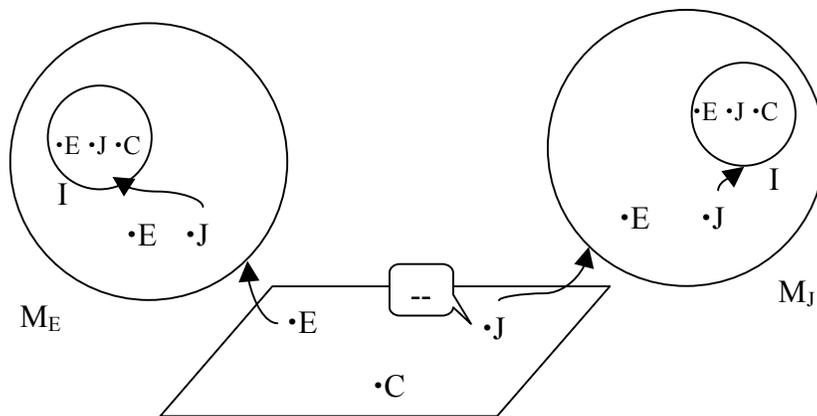

FIGURE 2. Reality space plus the discourse and intention spaces for the two signers.

3.3. COMMON GROUND SPACES. Finally, four different COMMON GROUND SPACES are used in the diagrams (see Figure 3). Each signer has a mental representation of the common ground they have with their spouse and also a mental representation of the common ground the couple has with the viewers. The common ground they have with each other is extensive and long-standing, though it increases with the making of this video. The making of the video itself increases the common ground between them, and some new information is shared in the video that does not seem to have been in their common ground previously.

      Since discrepancies can arise between the speakers as to what is in the common ground, each signer's representation will be shown in my diagrams, with separate spaces for their common ground with each other (CG_{EJ}—the common ground between Elaine and John) and the common ground with the viewer (CG_{EJC}—the common ground among Elaine, John, and the audience (represented by the camera)).
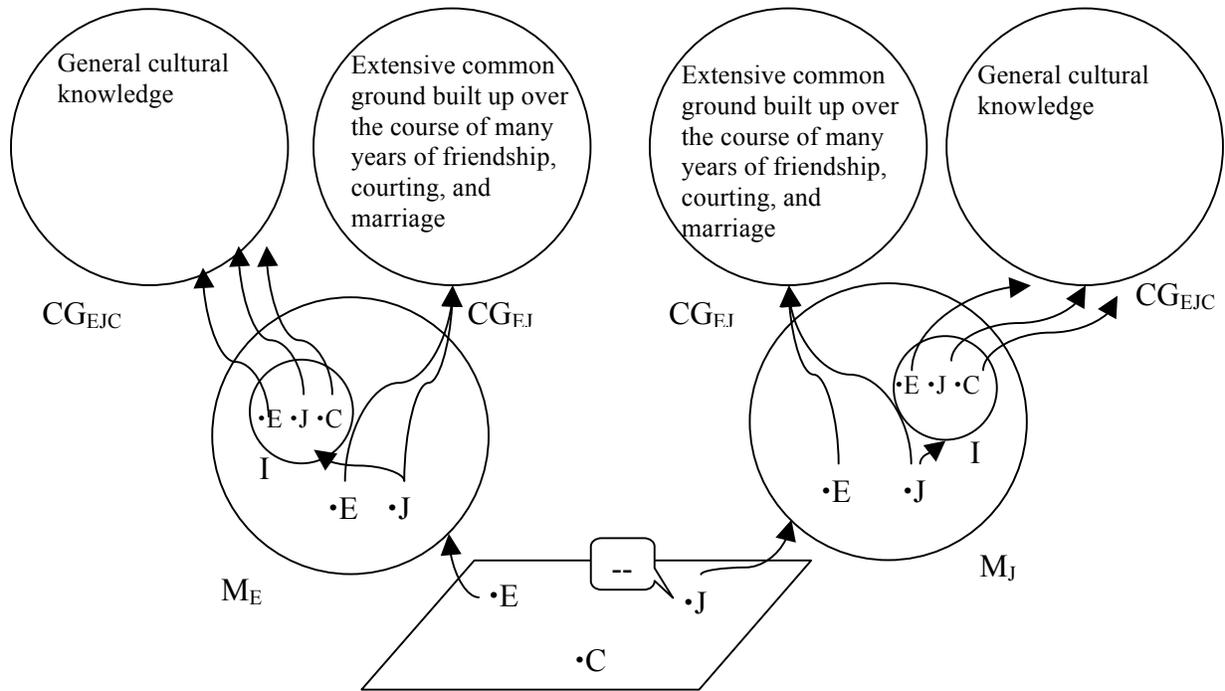
FIGURE 3. The mental space diagram at the beginning of the video.

3.4. ESTABLISHING COMMON GROUND: INTRODUCTIONS. The first two segments of the video involve the couple introducing themselves. First John Barton introduces himself, giving, as is traditional in Deaf introductions, the hearing status of his family and also naming the different places he's lived. These facts are added to the common ground without dispute. All this information seems to already be in the common ground between Elaine and John, given Elaine's attention and lack of surprise as he introduces himself. Figure 4 shows the status of the common ground as it exists after John's introduction. The information is the same in the representations of all four common grounds. While Elaine and John have considerably more than this in their common ground, that other knowledge is not in focus and most of it is not pertinent to the current joint project.

4. REPAIRS. In the course of the narrative a number of repairs occur, illustrating several different common-ground-based motivations.

4.1. GETTING THE IMPORTANT STUFF RIGHT: FAMILY INTRODUCTIONS. There is an interesting juxtaposition in this video because John and Elaine each introduce their children. John's introduction is split off as a separate segment of the video, and Elaine's introduction appears at the beginning of the long segment that is mostly taken up with the birth of their eldest child.

Lists of children are a structured piece of discourse in ASL. The non-dominant hand forms a LIST BUOY (Liddell 2003:223–242) with the number of fingers extended matching the number of children. The dominant hand then touches each finger in turn, with the top finger representing the oldest child and the bottom representing the youngest, and after touching the corresponding finger, gives information about that child. This information typically includes the child's name, name-sign, age, hearing status, gender, and any other information the parent

159

wishes to include. As we'll see in these introductions, there is a fair amount of leeway in the information included, but the form of the introduction is strongly present.
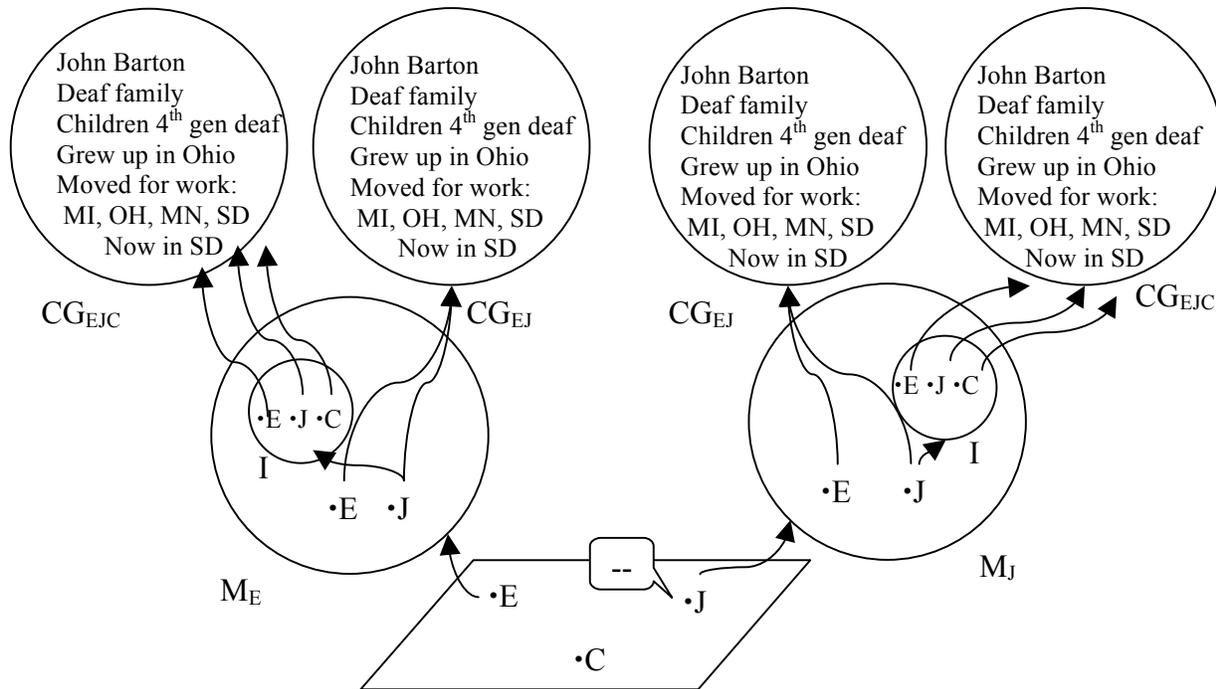

FIGURE 5. Common Ground after John's introduction.

John's introduction of the children is given in its entirety in 1. For the first two children he gives their ages, names, and name-signs, then when he gets to the third child he realizes he hasn't given their hearing statuses so he goes back to add them (and forgets to add it for the subsequent children). For the fourth child (line 6) he puts her hearing status where he had put the name-signs of the older children. Elaine senses the tension of his introduction not matching the expectation she has for the framework of these additions to CG_{EJC}, and she gives Ernestine's name-sign while John tells her hearing status. It isn't clear whether she is trying to prompt him or if she doesn't expect him to see. In any case, John isn't watching her and her contribution passes unaccepted and unacknowledged. However, we can see in Figure 5 the framework expected for the addition of new information to the common ground. Because he gave the first three children's information in the same order, Elaine (and the viewers) expected that he would continue in the same pattern. John, however, did not necessarily have the same expectation and in this case switched the order of the last two elements.

(1) JB: O-K 5-LIST-1 EIGHT NAME E-S-T-H-E-R ESTHER-NAME-SIGN[3]
       OK. Of our five children, the oldest is eight and is named Esther.

     JB: 5-LIST-2 AGE-SIX J-O-H-N-N-Y P-A-L SUMMARY NAME J-P
       Our second child is six years old. Johnny Pal--we call him JP for short.

160

→ JB: 5-LIST-3 OH 5-LIST-2 DEAF 5-LIST-1 DEAF
Our third child, oh wait, our second child is deaf, and our oldest child is deaf.

JB: 5-LIST-3 THREE AGE-FOUR 5-LIST-3 E-M-I-L-Y EMILY-NAME-SIGN
Our third child is three, four years old, Emily.

JB: 5-LIST-4 E- AGE-TWO TWO WILL AGE-THREE 2-MONTHS E-R-N-E-S-T-I-N-E
Our fourth child is E-, two years old, two, will be three in two months, Ernestine.

(0.5)

→ JB: [HEARING ONE] ERNESTINE-NAME-SIGN 5-LIST-4
That child is hearing--Ernestine.

→ EB:[ERNESTINE-NAME-SIGN]
Ernestine

JB: 5-LIST-5 J-A-S-O-N JASON-NAME-SIGN BOY AGE-ONE FOURTEEN MONTHS
Our youngest child is Jason, a boy. He's one year old, actually 14 months.
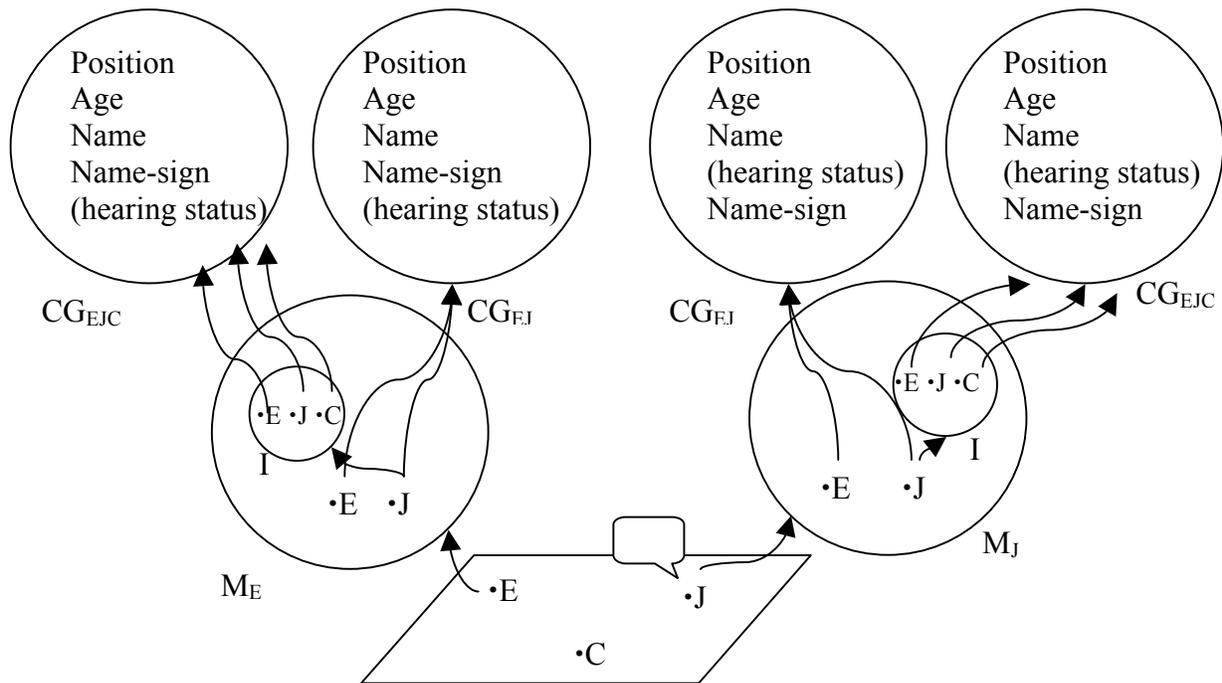
FIGURE 5. Expected form of introduction.

161

Elaine also introduces the children, but she does not need to give the information supplied by John, since that has already been placed in the CG_EJC. Her introduction, given in 2, is a different implementation of the listing structure, in that she discusses them in groups. She starts with the age range, then groups them by hearing status, then by gender.

(2) EB: TWO-OF-US HAVE FIVE CHILDREN
       'We have five children.'

EB: 5-LIST-1 EIGHT 5-LIST-1 5-LIST-5 AGE-ONE
       'The oldest is eight and the youngest is one.'

EB: NOW 5-LIST-ALL 5-LIST-1-2-3-5 DEAF 5-LIST-4 HEARING
       'Now, of all our children, numbers 1, 2, 3, and 5 are deaf, and number 4 is hearing.'

EB: 5-LIST-4 GIRL
       'She's a girl.'

→ EB: HAVE THREE GIRLS 5-LIST-1 5-LIST-2 5-LIST-4 TWO BOYS 5-LIST-3 5-LIST-5
       'We have three girls, numbers 1, 2, and 4, and two boys, numbers 3 and 5.'

JB: WHOA HAVE CLEAR MISTAKE
       'Wait a minute, we obviously have a mistake here!'

EB: [O-O-P-S]
       'Oops!'

→ JB: [5-LIST-2] BOY 5-LIST-4 5-LIST-5 BOY
       'Number 2 is a boy, and number 4, no 5, is a boy.'

JB: LOSE COUNT O-K
       'You lost count; that's ok.'

It's at this point that trouble appears, as she makes a mistake in saying which children are boys and he jumps in to repair the error. Other-repair is overwhelmingly initiated in the turn following the error (Schegloff, Jefferson, & Sacks 1977:367), and that is the position in which it occurs here. Even though there are no regular turns being employed at this point in the discourse, he initiates the repair at the end of her prosodic unit, the place where a turn would normally be taken in conversation. In line 7 of 2 Elaine acknowledges her error, but he is not looking and her 'oops' goes unaccepted, and she allows him to complete the repair. The common ground at the point just before he initiates repair is shown in Figure 6.

5 children
ages 1–8
1, 2, 3, 5 deaf
4 hearing
1, 3, 4 girls
2, 5 boys

5 children
ages 1–8
1, 2, 3, 5 deaf
4 hearing
1, 3, 4 girls
2, 5 boys

5 children
ages 1–8
1, 2, 3, 5 deaf
4 hearing
1, 3, 4 girls
2, 5 boys

5 children
ages 1–8
1, 2, 3, 5 deaf
4 hearing
1, 2, 4 girls
3, 5 boys

CG_{EJC} CG_{EJ} CG_{EJ} CG_{EJC}

·E ·J ·C   ·E ·J ·C

I   I

·E   ·J   ·E   ·J

M_E -- M_J

·E   ·J

·C
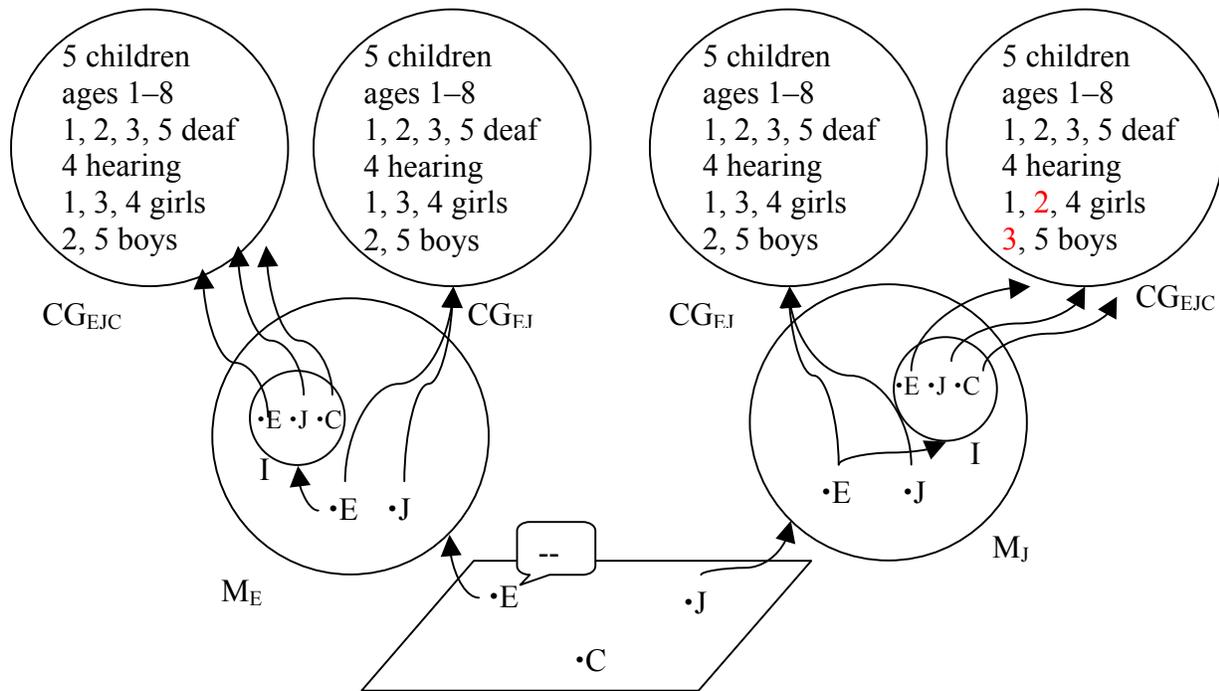
FIGURE 6. Common Ground at the point of repair.

As far as Elaine knows at this point she has communicated accurate information about her children. And there is no doubt that John believes Elaine knows which of her children are girls and which are boys, but he recognizes a discrepancy in the common ground with the unseen audience C, so he makes the repair.

It is interesting that each signer initiated some level of repair when the other was introducing the children. There is evidently some level of anxiety that this information be added properly to the common ground.

4.2. 'NO-FAULT' REPAIRS. Schegloff, Jefferson, and Sacks (1977:363) point out that sometimes 'repair [is] initiated with no apparent error,' such as replacing a word with a synonym or aborting a phrase and restarting with a different phrase, as in 3. Because of this, they say that 'it appears that nothing is, in principle, excludable from the class "repairable".' However, I propose that the trouble is not something that is present in speech, but rather arises in the mental representations of the common ground between the speakers or between the speakers and other listeners.

(3)   PRO-1 FINE
        that was fine with me

        PAR-- YOUNG COME-ON
        Par-- We were young and just said Oh, come on!

In 3, Elaine starts to sign PART but stops and changes to YOUNG. She may have been going to sign something along the lines of 'Part of the reason...' or something like that but

realized that that wasn't what she wanted to put into the common ground. This discrepancy led to a change in what she was signing.

4.3. TELLING A PARALLEL TRACK OF THE STORY. The most common situation where overlaps and interruptions occurred in this narrative were times when there were several things going on at the same time in the story, and each partner had a different part they want to emphasize. In 4 we see Elaine talking about getting ready for the induction of labor. She tells of her excitement, getting the injection, and then waiting. John jumps in on line 2 with a description of his readiness and the time passing while Elaine continues to emphasize the waiting. Then he starts talking about the Lamaze breathing, which had slipped from her focus in her memory of the event, as evidenced by her remark on line 10.

(4) EB: HAVE INJECTION [WAIT]
        'I got the injection, then we waited.'

→ JB: [READY DRESSED CLEAN] HANDS READY STRETCHING
                                'I was ready, got the scrubs on, scrubbed my hands.'

  JB: ENERGY READY
        'Feeling lots of energy, really ready'

  JB: MORNING BECOME AFTERNOON
        'Then morning turned into afternoon'

  JB: READY STRETCHING
        'We were still ready.'

→ EB: [PRO-1 WAIT]
        'I waited.'

  JB: [UNEXPECTED] #DO HAPPEN [WHAT WHAT NOTHING WHAT]
        'Unexpectedly, nothing happened.'

  EB: [WAIT TWENTY H-O-U-R-S]
                                        'I waited twenty hours!'

→ JB: NOT-YET TWO-INDIVIDUALS-FACING-EACH-OTHER [BREATHE]
        'Nothing yet, but we started breathing.'

→ EB: [PRO-JB BREATHING RIGHT PRO-JB]
                                            'Oh yeah, breathing, right'

  JB: PRO-EB NOT-YET++ WAIT COME-ON WRONG
        'You were saying, "Not yet, come on. What's wrong?"'

In this situation it is not so much a discrepancy of content among the common ground mental spaces that is an issue, but rather a discrepancy of focus, as illustrated in Figure 7. I have highlighted the focus information for each speaker and placed Elaine's memory of breathing in parentheses to show its very low focus level. The CG$_{EJC}$ spaces are shown as at the end of 4.
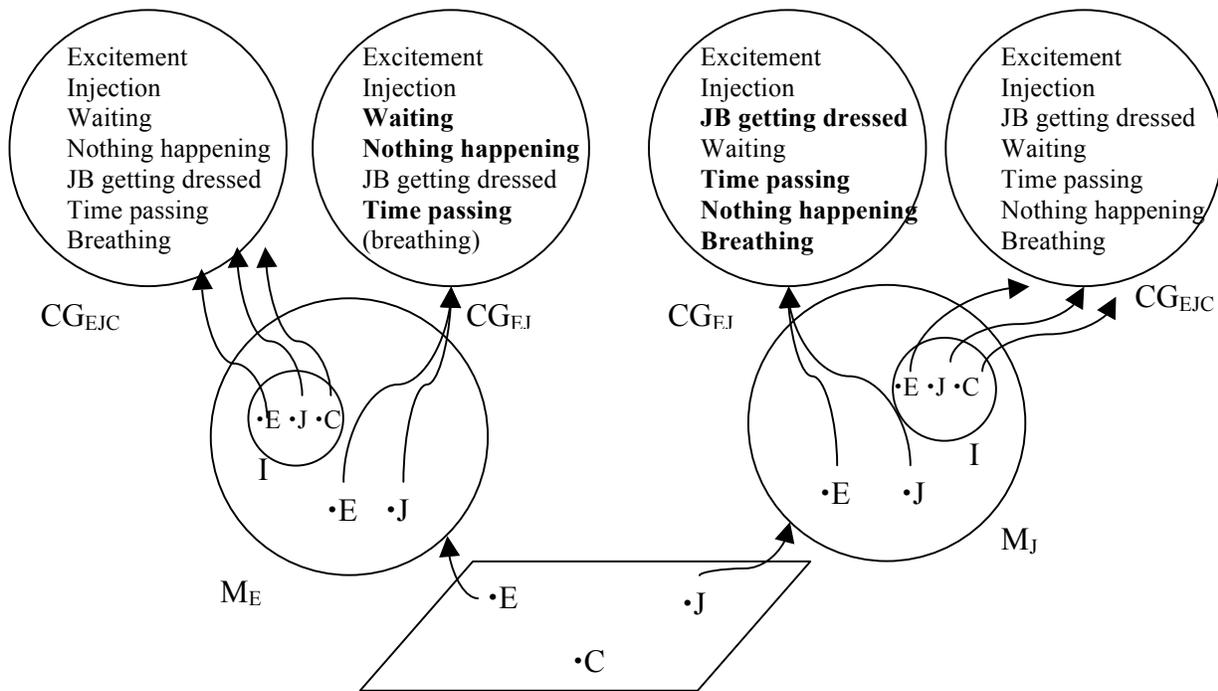

FIGURE 7. Different focus information.

In this passage, no speech error or misstatement of fact has been made. However, there is tension in the common ground that needs to be resolved. This resolution is achieved by the spontaneous offering of additional details, either in turn or by co-speaking or interrupting.

5. CONCLUSION. There are many configurations in which common grounds can exhibit discrepancies. In the narrative I have looked at here we have seen discrepancies in expected frameworks, in the common ground shared with the viewer but not with each other, in the common ground shared with each other, and in focus elements. Each time repair has been initiated by the party who experienced a discrepancy between their representations of CG$_{EJ}$ and CG$_{EJC}$ at the time the discrepancy became apparent. The tension involved in having differences between the two common grounds is resolved through repair.

The ability to represent common grounds in a mental space framework allows us to see these discrepancies at the cognitive level where they cause tension that the speaker feels the need to resolve. Past discussions of repair have looked only at the surface manifestations and have been unable to account for some repairs where the trouble is not an actual error. However, this model shows the motivation behind all repair as the desire to resolve discrepancies in the common ground.

REFERENCES

CLARK, HERBERT. 1996. *Using language*. Cambridge: Cambridge University Press.
CLARK, HERBERT, and EDWARD SCHAEFER. 1989. Contributing to discourse. *Cognitive Science* 13.259–294.
FAUCONNIER, GILLES. 1994. *Mental spaces: Aspects of meaning construction in natural language*. Cambridge: Cambridge University Press.
KONEFAL, JOANNE A., and JOANN FOKES. 1984. Linguistic analysis of children's conversational repairs. *Journal of Psycholinguistic Research* 13.1–11.
LIDDELL, SCOTT K. 2003. *Grammar, gesture, and meaning in American Sign Language*. New York: Cambridge University Press.
LIEBSCHER, GRIT, and JENNIFER DAILEY-O'CAIN. 2003. Conversational repair as a role-defining mechanism in classroom interaction. *The Modern Language Journal* 87.375–390. Online: doi:10.1111/1540-4781.00196.
RSA REGION V PROJECT. 2003. All in due time: Perspectives on childbirth from Deaf parents. *Interpreting in Healthcare Settings*. Online: http://www.medicalinterpreting.org/Interpreting/CDROMS/DueTime/index.html.
SALONEN, TUULI, and MINNA LAAKSO. 2009. Self-repair of speech by four-year-old Finnish children. *Journal of Child Language* 36.855–882. Online: doi:10.1017/S0305000908009240.
SCHEGLOFF, EMANUEL A. 2000. When "others" initiate repair. *Applied Linguistics* 21.205 –243. Online: doi:10.1093/applin/21.2.205.
SCHEGLOFF, EMANUEL A.; GAIL JEFFERSON; and HARVEY SACKS. 1977. The preference for self-correction in the organization of repair in conversation. *Language* 53.361–382. Online: doi:10.2307/413107.
SVENNEVIG, JAN. 2008. Trying the easiest solution first in other-initiation of repair. *Journal of Pragmatics* 40.333–348. Online: doi:10.1016/j.pragma.2007.11.007.

[1] I follow the standard convention of using deaf to refer to people who have the physical condition of deafness and Deaf to indicate those who are culturally Deaf.

[2] The names of the speakers have been changed to preserve their privacy.

[3] GLOSSING CONVENTIONS:

NAME	An uppercase English word identifies a single ASL sign.
GROW-UP	Uppercase English words separated by hyphens also represent a single sign.
O-H-I-O	Single letters separated by hyphens indicate fingerspelled words.
SAME$^{\rightarrow \text{JB}}$	$^{\rightarrow \text{x}}$ indicates that the sign is directed toward entity x, in this case JB.
5-LIST-2	The first number indicates the number of fingers in a list buoy, the second the finger selected
NOT-YET++	+ indicates a reduplication of a sign, which can signify plurality of nouns or intensification.