Proceedings of the Twenty-Ninth

Western Conference on Linguistics

Volume Twenty-Three

WECOL 2017

Edited by
Michal Temkin Martinez
Gail Shuck
Tim Thornes

Held at:
Boise State University
October 20 – 21, 2017

Department of Linguistics
California State University, Fresno

Contents

Towards an Analysis of DP Structure in Telugu

Akshay Aitha
University of California, Berkeley

1 Introduction

Telugu (tel, South-Central Dravidian) does not feature definite or indefinite articles. The bare nominal argument can have either a definite or an indefinite interpretation depending on context.

(1) kukka naDus-tun-di
dog walk-PROG-3SG.NHUM
(A/The) dog is walking.

I posit, contra claims made by Boskovic (2006) that languages without overt articles do not have DPs, that Telugu does in fact show evidence for a DP layer. To do this, I will show that the bolded phi-agreement morphemes[1] in the below examples are nominative-case-marking clitics of category D in whose specifier position are found Telugu pronouns[2], and that the accusative and dative case markers are also of this category. I support this conclusion using facts about intra-nominal agreement, the distribution of the morphemes in question, and general word order in Telugu complex subjects.

(2) a. {meemu padi-mandi pilla-la=**mu**}$_{DP}$ idi cees-tunnaa-mu
1PL.EXCL.PRO ten-HUM child-PL-1PL.NOM this do-PROG-1PL
We ten children are doing this.

b. {neenu okkaDi=**ni**}$_{DP}$ idi cees-tunnaa-nu
1SG.PRO alone-1SG.NOM this do-PROG-1SG
I alone am doing this.

c. {nuwwu okkaDi=**wi**}$_{DP}$ idi cees-tunnaa-wu
2SG.PRO alone=2SG.NOM this do-PROG-2SG
You (sg) alone are doing this.

In doing this, I also argue for the DP structure shown below:

(3) {meemu padi-mandi pilla-la=**mu**}
1PL.EXCL.PRO ten-HUM child-PL-1PL.NOM
We ten children

*I would like to acknowledge the comments given to me by the audience at WECOL 2017 and at the weekly Syntax & Semantics Circle in my home department at UC Berkeley.

[1]Besides first-person plural, first-person singular, and second-person singular, the rest of the paradigm is null. For example:

{miiru padi-mandi pilla-lu-ø} idi cees-tunnaa-ru
2PL.PRO.NOM ten-HUM child-PL-2PL.NOM this do-PROG-2PL
You ten children are doing this.

[2]For sentences like (1) which do not have any pronouns or any concord-like suffixes, I posit that D is null and has default features, and that its specifier position is empty.

Proceedings of *WECOL 2017*

classified by UNESCO in Moseley (2010). Lewis, Simons and Fennig (2017) rate Cherokee as a 6b on their endangered scale, meaning that although the language is spoken intergenerationally, the language is losing speakers. However, Lewis, Simons, and Fennig's (2017) rating does not take into account actual, recent work done in the community to assess the language situation. Scancarelli (1986) and Montgomery-Anderson (2015), both active in Cherokee language work, have found that there are consistently no speakers under the age of 40 and that the language is not being transmitted in the home. Therefore, a better place to rate Cherokee would be at a 7 on the GIDS Typology (Fishman, 1991), placing Cherokee as a more endangered language than what Lewis, Simons, and Fennig (2017) assert.

2.2 *Current revitalization efforts*

The rate of language attrition might seem shocking to some at first due to the immense variety of resources available to language learners, from online and technological resources to plenty of printed materials and in-person classes. The Cherokee Nation offers online language classes where students can see a live video of a fluent speaker going through a PowerPoint presentation of ten phrases, as well as numerous downloadable resources and an easy-to-use word list (although it is not extensive). The United Keetowah Band of Cherokees has documents with traditional stories available for free. A couple of mobile apps exist, as well as a Cherokee keyboard for smart phones and Microsoft software. Another online class, created by speakers, teaches grammar and tone through uploaded videos and a textbook. For printed materials, a reference grammar has been published, as well as a Cherokee-English dictionary. The Cherokee Phoenix is still circulating, as well as a magazine called *Anadisgoi*, with both featuring the Cherokee syllabary in addition to English. In-person classes range from immersion schools (offered by Cherokee Nation and the United Keetowah Band) to college level courses at Northeastern in Tahlequah, Oklahoma, to immersion camps hosted multi-annually and other community-level classes. Recently, the Cherokee Nation has also developed a master-apprenticeship program, although spaces are extremely limited and competitive to get into.

3.0 Background

In reality, this research began not from a question but rather an observance, that "heritage language learners of Indigenous languages who do not have access to a community of practice will need different linguistic resources." To adequately address this observation, however, it is first necessary to understand what the best practices and common methodologies currently are for learning Indigenous languages, as well as who heritage language learners are and how these learners can achieve their learning goals.

3.1 *Indigenous heritage language learners*

It should come as no surprise that Indigenous heritage language learners come in all varieties, from younger to older, traditional to non-traditional, and in-community to peripheral, just to name a few. As displacement is a primary factor in language attrition (Crystal, 2000; Harrison, 2007; McCarty & Zepeda, 2010), it is probable that most Indigenous communities of practice have individuals who live away from the geographical area of practice but still identify as a community member. Unfortunately for these peripheral heritage language learners, though, they have grown up in a dominant culture which has excluded them from their heritage and, as LaFramboise et al. (1993) asserts, has required Native-identifying individuals to pass as "culturally competent" (p. 396) by adhering to all socially sanctioned behaviors, having a strong identity based in the dominant culture, knowing of all beliefs and values of the dominant culture, using the approved language of the culture, and actively being a full member of the community. Essentially, the dominant culture not only removes Native-identifying individuals from their respective communities, but it further bars them from any identification factors that they could turn to in order to project their heritage identity. Many who still harbor a connection to their Native heritage, however, turn to language learning in adulthood as a gateway to the community and to unlocking their self-ascribed identity (Hale & Hinton, 2001; Nicholas, 2009; Peter, 2014; Kickham, 2015; White, 2015).

Therefore, Native-identifying individuals turning to language learning are, in fact, desiring cultural and communicative competency (i.e., the ability to produce language in a meaningful context) instead of linguistic competency (i.e, the ability to generate a grammatically correct phrase or sentence, regardless of context or meaning). For Indigenous heritage language learners, the importance of language learning rests

not on the ability to formulate grammatical sentences regardless of context, but on being able to have a meaningful conversation, to respond to unique cultural variables, and to be a successful member of society by adhering to the parameters of inclusion imposed by language. With this motive, distance resources can (and often do) lose their function, since most erase the culture and community – the two connections that Indigenous heritage language learners seek most.

3.2 *Common methods for learning Indigenous languages*

As endangered languages gain more spotlight in linguistics and anthropology, research on language revitalization, reclamation, documentation, and maintenance (and, specifically, on best practices for teaching and learning Indigenous languages) is becoming more prominent. Various sources agree that methods that immerse learners solely in the target language are the most effective; immersion-style schools and language nests continually boast reversal of language loss, first in Hawaii and New Zealand, and now in North America (Hale & Hinton, 2001). Additionally, master-apprentice programs are being implemented in multiple Indigenous communities across the globe to much success (Mentor-Apprentice Program, 2008; Olawsky, 2013; National Breath of Life Archival Institute for Indigenous Languages, 2016). Furthermore, language classes put on by the community or by educational institutions (such as universities and high schools) attract heritage language learners and help produce speakers. These three methods flourish for numerous reasons, but most notably, they are all in-person methods where the learner has direct, physical access to a community of practice and/or a language teacher. As of now, research in language revitalization has yet to delve into methods for distance learners or the use of technology in transmitting Indigenous heritage languages.

Although in-person attendance in immersion-style programs is the ideal learning environment for Indigenous languages, it requires learners to have physical proximity to a community of practice (Hale & Hinton, 2001; Mentor-Apprentice Program, 2008; Olawsky, 2013; National Breath of Life Archival Institute for Indigenous Languages, 2016). As distance learners, displaced community members naturally do not have the option to take in-person classes, but must instead rely on distance materials, such as books, internet downloads, or online classes. However, the majority of these resources fail to incorporate interaction between members of a community of practice. Distance heritage language learners, then, have little to no resources that will satisfy their motives for turning to language learning; since they cannot attend in-person programs and most distance materials lack a connection to a community, distance heritage language learners cannot gain cultural or communicative competency like they desire.

4.0 Methods

Ethnographic accounts offer rich glimpses into life, especially into worlds gone unnoticed. Specifically for Native scholars in linguistic anthropology, presenting one's experience and research via auto-ethnography is becoming increasingly more accepted, a shift that gives credence to perhaps otherwise unspoken views. Thus, as a Native-identifying individual who is a member of this study's target demographic, I derive data from my auto-ethnographic account of growing up as a peripheral heritage language learner of Cherokee. In an attempt to include more voices, this research also includes personal communication from two older males and one older female living in geographically distant places. All four of the views presented in this paper are from peripheral heritage language learners of Cherokee who have interacted with a significant amount of the available in-person and distance linguistic resources.

5.0 Data and discussion: Resources

To begin, this paper will discuss which elements of available linguistic resources peripheral heritage language learners of Cherokee view as most successful in meeting their goals. Presented accounts identify having access to a Cherokee community of practice and achieving cultural and communicative competency as most important, even more so than linguistic competency.

5.1 *My account*

My language learning journey has taken me through all manner of resources, from online material to printed sources like dictionaries, story books, and a reference grammar. Thinking at first that I would learn

Cherokee how I have learned other languages, I first turned to using mobile applications and books, which worked well for about a month, but I quickly found myself deeply frustrated with my inability to parse out vocabulary, and I began to yearn for a more profound experience with the language (though I could not identify yet what this was or would look like). I then discovered online classes provided by the Cherokee Nation, and, excited with the prospect of interacting with a speaker and other beginning learners in real-time, I signed up. Due to various reasons, however, the classes did no more for my Cherokee than my previous independent studying. I could not internalize anything greater than *osiyo* ('hello') and, though the teacher was brilliant, I could not connect what he was saying to the language. Not giving up, though, I flew to the Cherokee Nation in Tahlequah, Oklahoma, to visit the teacher who leads the classes. After connecting with the teacher in-person and establishing myself as a Cherokee heritage language learner, the online classes held new meaning for me; instead of being a list of foreign vocabulary, the context behind the language came alive, and I finally began internalizing the language. Linguistic competency was not what I was seeking, and, truthfully, not even possible for me until I demonstrated that I was a community member and developed communicative and cultural competency. The most important aspect of linguistic resources for me, then, has proven to be access to a community of practice and an understanding of the culture behind the language.

5.2 *T's account*

T is an older male who has interacted with numerous resources, but prefers to take online classes – one being the online class that Cherokee Nation offers with a speaker in real-time, and the other being an independent study class that uses a grammar textbook (with the description that it is not to be used to learn conversational Cherokee). T named the Cherokee Nation class with a speaker as a top resource because, "It was great for connecting to a community of learners." Furthermore, T identified a major flaw with the grammar class that was self-paced, saying that there was "no interaction with live humans."

These online resources are arguably the best from a linguistic and student's standpoint. Offering a live fluent speaker, Cherokee Nation's online class imitates an in-person language class (though students can only ask questions and interact with other students via sending a chat message to the class). Additionally, the grammar intensive online class offers rich linguistic knowledge of Cherokee. Combined, both resources should teach students conversational phrases and how to build their own sentences. However, notice that T does not point to the best part of Cherokee Nation's class as having a live speaker to go over phrases or to answer grammatical questions, or even any reason relating to fluency or linguistic elements. Instead, he focuses on how the class creates a bridge to other learners. He further clarifies a drawback in the other class, stating that it has no interaction with live humans. Although the grammar-intensive class can build linguistic competency, it cannot provide community access. His primary interest in classes entirely circumvents linguistic competency and hones in on access to a community of practice.

5.3 *S's account*

S is an older female who primarily uses CDs and Cherokee Nation's online classes to learn Cherokee. She said, "If you can find a person to communicate on a regular basis, it is even more helpful. I am in an area where it is hard to find a person I can just talk to and learn from." Learners use classes and CDs to guide their pronunciation in order to speak Cherokee. However, even when using resources that promote pronunciation and conversations, there is still a gap in resource utility. With the resources that S uses, she can become linguistically competent by learning phrases through the online classes and practicing "correct" pronunciation with the CDs. However, for her, regular communication with a speaker at any level (notice she does not specify a fluent speaker or other learner) is what is most important. She also specifically identifies herself as a peripheral language learner with no immediate access to a community of practice, acknowledging that as a learner, she is at a disadvantage from others with easier access to speakers. In S's case, it is apparent that even when resources can be used for vocabulary learning and pronunciation, the distance between learner and community of practice is still a tremendous issue, and the importance of communicative competence supersedes that of linguistic competence.

5.4 *J's account*

Lastly, J is an older male who has spent many years learning Cherokee by using every resource imaginable. A peripheral language learner who has worked tirelessly to achieve linguistic competency, J is

b. *naannagaaru pilav-anDi

(21) a. kukka=**ni** pilav-anDi
Father=ACC call-2PL.IMP
Please call the dog.
b. *kukka pilav-anDi

When the object of the verb is inanimate (as in *kaafii* 'coffee'), the accusative case marker *-ni* is optional. When it is animate, leaving out the accusative marker is ungrammatical. This is a DOM (differential object marking) system, in which animate objects with more agency (and thus more likely to be misparsed as a subject) require the accusative marker, whereas inanimate objects do not require the accusative marker. From this, we make the generalization that the case marker position is sensitive to the human-ness feature. To make the claim that the pronoun position is also sensitive to that feature, we must look to the demonstrative paradigm in Telugu, since the use of pronouns already implies human-ness.

(22) aa muuDu pustakaa-lu
those.NHUM three.NHUM book-PL
those books

(23)

	Human	**Non-Human**
Proximal	wiiLLu	ii
Distal	waaLLu	aa

Demonstratives in Telugu compete with and occur in the same linear position as pronouns, which can be taken as evidence that they occur in the same syntactic position. The paradigm shows that different demonstratives are used for human and non-human nouns, showing the sensitivity of this position to the human-ness feature. Looking down into the NumP, numerals are sensitive to number/human-ness features because alternation based on human-ness only occurs in the plural; the paradigm has been summarized here[5]:

(24)

	oka - 'one'	**muuDu - 'three'**	**padi - 'ten'**
Human	oka	mug-guru	padi-mandi
Non-Human	oka	muuDu	padi

We note that neither numerals, nouns, adjectives, nor plural markers show any sensitivity to person or case. Not only does this description of feature sensitivity solidify the claim that pronouns/demonstratives and case markers exist at the same level of structure, it is also good evidence for that layer's high position in the nominal. Since Agree transfers features from low in the tree to high, it would follow that a given feature can only be found where it is generated and above, and that the layers sensitive to the most features should be the highest.

5 Piecing Together the Geometry of the DP

To determine the structure of the DP below the D projection, we can first look to a telling piece of the distribution of the plural marker:

(25) anna tammuL-Lu
older.brother younger.brother-PL
'brother' (lit. older brothers and younger brothers)

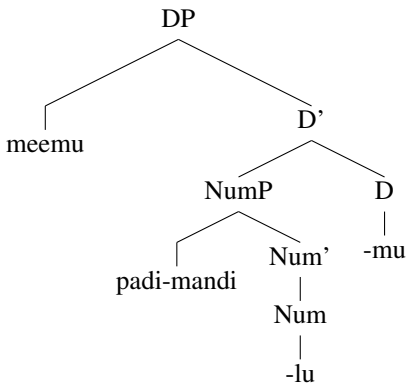
The plural marker *-lu* applies to both elements in a coordination without having to attach to each one. A very similar phenomenon in Turkish (Hankamer 2012) has been called an example of suspended affixation; Hankamer concludes that the Turkish affix in question behaves this way because it is in fact a head which has an NP as its complement, allowing the plurality to scope over multiple coordinated elements. The suspended affixation-like examples lead to an easy story for the Telugu plural marker as a

[5] *-mandi* is the human suffix for numerals 8 or greater; from 2 to 7 it's *-guru* with stem allomorphy.

Num head which occurs a layer above the N layer at which this coordination occurs. A further detail of the distribution of *-lu* is illustrated by this example, a repeat from earlier:

(26) {meemu padi-mandi=**mi**} idi cees-tunnaa-mu
1PL.EXCL.PRO ten-HUM=1PL.NOM this do-PROG-1PL
We three are doing this.

Given the proposed analysis, the proper structure for the subject of this sentence would be as follows:



Since the numeral, which is in Spec,NumP, is overt, a Num head must also occur in the structure. To explain the fact that there is no overt *-lu* in this sentence, the generalization can be made that the suffix requires material below it (i. e. a noun) to support it. From there, two possible analyses for how the noun supports the plural suffix become evident - (1) a head-movement story in which N moves to Num in the syntax and (2) an affix-hopping story in which *-lu* hops down to N in the morphology. Deciding between them, especially in a head-final language like Telugu where head movement would not go past specifiers, may be difficult. It suffices to say that the theory provides (multiple) parsimonious ways to explain the distribution of *-lu* assuming the proposed structure, showing that placing numerals higher than both nouns and the plural marker is desirable.

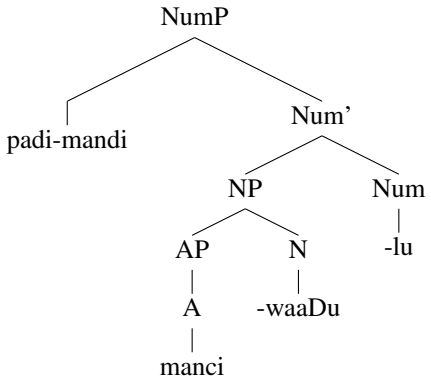
Support for the structure of NumP can be found by synthesizing work done by Greenberg (1977) and Watanabe (2006). Greenberg makes the typological generalization that languages with numeral classifiers can never have obligatory plural marking. One interpretation of this is that plural markers and classifiers are of the same category cross-linguistically. Obligatory plural marking would mean that classifiers could not fit anywhere in the syntax, because number markers always take up their spot. Watanabe, in a paper about Japanese DP structure, argues that Japanese numerals appear in the specifier position of classifiers. Thus, if the Telugu plural marker is of the same category, existing literature provides some support for the claim that Telugu numerals occur in the specifier position of the plural element. General evidence for this configuration is given by the word order in nominals such as *meemu padi-mandi pilla-la-mu*, in which the numeral and plural marker appear on opposite sides of the nominal root *pilla*, showing that the layer of structure that they define contains the layer at which the nominal root occurs. Also, since Telugu is a head-final language, assuming that specifiers always attach to the left generates the expected word order.

An additional claim that can be made is that Telugu numerals and Telugu adjectives occur at separate levels of structure.

(27) {meemu manci-waaLLa=**mu**} idi cees-tunnaa-mu
1PL.EXCL.PRO good-NOMZ.HUM=1PL.NOM this do-PROG-1PL
We good ones are doing this.

When using an adjective without a lexical noun, a nominalizer (exponing human or non-human features) appears between the adjective and the plural marker, where the nominal stem would usually occur. This is in contrast to what occurs when a numeral is used without a lexical noun, as in (26); in that case, there is no nominalizer required. This difference in behavior can be analyzed as the N projection (at which the nominalizer occurs) being higher than the adjective, but lower than the numeral, and by saying that in sentences such as the one in (26), it's the NP which is being elided. Thus, we need a nominalizer with an adjective because AP cannot Merge with Num, but since numerals are higher than nouns (as opposed to

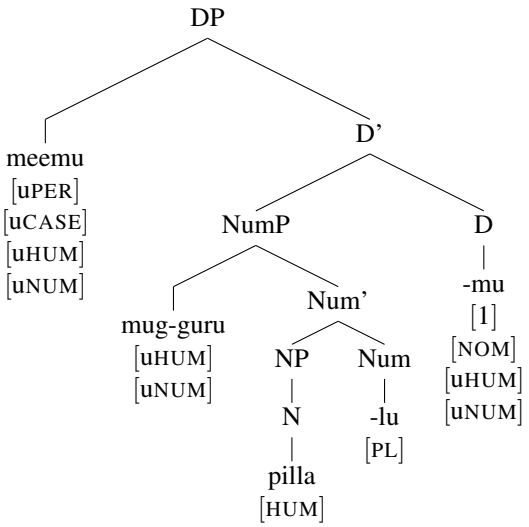
being in the adjective position), they do not require the support of a nominalizer to occur without a noun. To illustrate this, the structure for the subject in (27) is given below.



6 Syntactic Derivation of an Example Sentence

Given the analysis so far, we can assign features to nodes and from there determine the probe structure required to explain all the feature sensitivities described in previous sections. The following figure is the tree from (3) with all features and probes indicated at their base-generated positions.

(28)



The justification for this configuration is better illustrated in the context of the table below, which summarizes the conclusions of §4.

(29)

Element	Sensitivity
#	HUM, NUM
D	PERS, CASE, HUM, NUM
Pro	PERS, CASE, HUM, NUM

Considering their semantic contributions and what we have established already, it would stand to reason to have the N head carry the human-ness feature, the Num head carry the plural feature, and the D head carry the case feature. This is borne out by the fact that nothing below Num shows sensitivity to plurality, the fact that nothing below D shows sensitivity to case, and the fact that both layers above N show sensitivity to human-ness. The person feature, as has been shown, only triggers morphological alternations in the D layer. Thus, we say that the person feature is generated at the D head. The same applies to the case feature; I will not address the question of case assignment in this paper, but I assume that it gets assigned to D.

To decide where and what probes are needed, we start from the bottom of the tree and insert uninterpreted features as required. The lowest element in the structure which shows allomorphy based on

agreement is the numeral. We call Spec,NumP a probe with the features [uNUM] and [uHUM]. Next, D requires a probe which looks for the same features, so that it can get up to its full set of four features. Finally, we have a probe on the pronoun which includes uninterpretable features for person, case, human-ness, and number, because the pronoun is sensitive to all four. Given this configuration, we can explain all feature sensitivities using a basic implementation of Agree. For ease of comprehension, the information from the tree above is repeated in the following table:

(30)

Head	**Inherent**	**Agreement**
N	HUM	N/A
Num	PL	N/A
#	N/A	HUM, PL
D	1, NOM	HUM, PL
Pro	N/A	1, NOM, HUM, PL

The final step of the derivation is the insertion of the morphology. It is here that we encounter a small problem that arises in DPs that consist of only the DP layer without overt NumP or NP:

(31) waaDu maa=**ku** adi icc-aa-Du
3SG.MALE.PRO.NOM 1PL.PRO=DAT that give-PST-3SG.MALE
He gave that to us.

(32) waadu mammal=**ni** cuus-aa-Du
3SG.MALE.PRO.NOM 1PL.PRO=ACC that see-PST-3SG.MALE
He saw us.

(33) meemu caduvu-tunnaa-mu
1PL.EXCL.PRO.NOM read-PROG-1PL
We are reading.

We see that in the dative and accusative, we get the expected pattern; the case marker is cliticized to the bare pronoun and nothing else. However, in the nominative, we only get the nominative form of the pronoun without the expected nominative case marker *-mu* following it. I analyze this as stemming from a version of the Repeated Morph Constraint; since *meemu* already ends with /mu/, adding a clitic with the same form (giving **meemu-mu*) is not preferred.

7 Concord in Numerals?

Sentences like (34)-(36), which look like case and person concord in numerals, were judged grammatical by my consultants. This would be a counterexample to a generalization made by Baker (2008) that person features are not subject to concord. However, there do appear to be subtle meaning differences when the case marker is placed on a non-final element of the DP. For example, in the third example above, having *-mu* on both the numeral *iddaru* and the plural noun *pilla-lu* gives the interpretation of having two DPs meaning 'we two' and 'children', where the second is predicated of the first. This is also the case for the accusative and the dative when placing the *-ni* and *-ki* on both a numeral and the noun it counts. Furthermore, doing this with the accusative and the dative results in, according to my consultant, prosodically 'difficult' sentences.

(34) waaDu {maa iddari=**ki**} {pilla-la=**ki**} adi icc-aa-Du
3SG.MALE.PRO.NOM 1PL.PRO two.HUM=DAT child-PL=DAT that give-PST-3SG.MALE
He gave that to us two, who are children.

(35) waaDu {maa iddari=**ni**} {pilla-la=**ni**} cuus-aa-Du
3SG.MALE.PRO.NOM 1PL.PRO two.HUM=ACC child-PL=ACC see-PST-3SG.MALE
He saw us two, who are children.

(36) {meemu iddara=**mu**} {pilla-la=**mu**} caduvu-tunnaa-mu
1PL.EXCL.PRO.NOM two.HUM=1PL.NOM child-PL=1PL.NOM read-PROG-1PL
We two, who are children, are reading.

I posit that what is happening here is that every instance of a case marker, as expected, marks its own DP. Thus in sentences like those above, the speaker pronounces two DPs in a single argument position of the main verb, and the interpretation of such a construction is as an appositive. It stands to reason, also, that DPs track some prosodic unit, from which follows the reported difficulty in pronouncing sentences like the first two in this section. This is also very likely connected to the agreement which occurs in nominal predication in Telugu, as shown in (37).

(37) a. meemu sTuDenTu-la=**mu**
1PL.EXCL.PRO.NOM student-PL=1PL.NOM
We are students.

b. neenu Akshay=**nu**
1SG.PRO Akshay=1SG.NOM
I am Akshay.

c. nuvvu Raj=**vu**
2SG.PRO Raj=2SG.NOM
You are Raj.

Because they both feature a predicational meaning, it is reasonable to say that the same structure has been invoked in the complex DPs of (34)-(36). Furthermore, this may be good evidence for the agreement morphemes in nominal predication actually being case markers. This would put Telugu in a class of languages including Latin (Wheelock 2011), in which nominal predication results in both the argument and the predicate receiving the same case assignment. The fact that both this agreement and the set of nominative case markers are the same provides more credence to this hypothesis. This analysis recalls work done by Pelletier (1993) on a similar set of Telugu data, which analyzes all instances of *-mu* as apposition and further attempts to give a single analysis of verbal agreement, the sort of intra-nominal agreement which is the main focus of this paper, and the agreement in nominal predication introduced above. Mine is a little less ambitious; nominals with only one case marker do not appear to have any predicational sense to them, and the nominal paradigm is different from the verbal one[6], allowing me to argue that there are three different phenomena occurring.

8 Further Directions

A lot of the constructions I elicited for this project are very uncommonly used in Telugu, so the grammaticality judgements I received were often jumbled and contradictory. Though I believe I've been able to piece together a set of coherent facts that are true for at least some speakers, a wider survey and elicitation of a lot more subjects could lead to new insights and more information about variation within this part of the language.

More narrowly, this data brings up interesting questions about probes and case-assignment. If one accepts my analysis, the question of why it is only nominative case-marking that shows person- and number-sensitivity arises. I have analyzed the other case markers as covertly agreeing for those features without exponing them, because such an analysis presented no serious problems for my claims and because it greatly simplified what I had to say about the probes within DP, but an analysis involving different probes for different case assignments may be worth looking into.

9 Conclusion

It has been shown that Telugu, while lacking overt articles, still features a high projection in its nominals which includes its case markers, its pronouns, and its demonstratives. While this projection does not look exactly like a canonical DP, it performs enough of the same functions to say that it is good evidence that Telugu fits a category of languages that Boskovic claims should not exist: Those languages without articles but with DP. My proposal, then, gives credence to the assumption made in most literature that nominals are DPs and that even nominals without overt D feature this layer in the syntax.

[6]In the verbal paradigm, *-ru* marks 2nd/3rd-person plural, but in the nominal paradigm this slot is null.

This has been shown through an analysis of the distribution of a set of morphemes occurring in Telugu complex nominals which match the pronouns in person, number, human-ness, and case. By first showing that they are clitics and then showing how they pattern very similarly to the Telugu accusative and dative case markers, they are analyzed as overt heads of a high projection which marks nominative case as well as (sometimes) phi-features. Furthermore, articulating the feature sensitivities of the various elements of the Telugu nominals provides evidence, along with support from literature, that pronouns and demonstratives occur in the specifier of this posited projection.

Besides articulating the structure and position of this projection (which is called DP), a proposal for lower parts of the structure, namely NumP and NP, have also been posited.

I hope that the facts and analysis presented here can contribute to the larger theoretical goals of modeling cross-linguistic variation in nominal structure and concord.

10 Acknowledgements

While many of the sentences here were constructed by me, I would like to thank all three of my consultants for providing important grammaticality judgements on said sentences while clarifying and confirming many of the observations made by me or by the scholars I drew from for this project. Furthermore, I would like to thank Professors Peter Jenks and Line Mikkelsen for providing important input and guidance throughout this project, both through one-on-one collaboration and through the medium of the graduate syntax course I took in Spring 2017. The organizers of WECOL 2017, too, deserve thanks for giving me the opportunity to present and publish my research. Finally, I would like to thank the Syntax & Semantics Circle at Berkeley both for inviting me to give a practice talk and for giving me helpful, constructive criticism.

References

BOSKOVIC, ZELJKO. 2006. What will you have, DP or NP? In *Proceedings of NELS 37*, 101-114.

GIUSTI, GIULIANA. 2002. The functional structure of noun phrases: A bare phrase structure approach. In *Functional Structure in DP and IP*, ed. G. Cinque, 54-90. Oxford: Oxford University Press.

POSTAL, PAUL. 1966. On So-Called 'Pronouns' in English. In *Monograph Series on Languages and Linguistics, vol. 19*, 177-206.

NORRIS, MARK. 2011. Towards an Analysis of Concord (In Icelandic). In *Proceedings of the 29th West Coast Conference on Formal Linguistics*, 205-213.

ZWICKY, ARNOLD, PULLUM, GEOFFREY. 1983. Cliticization vs. Inflection: English N'T. In *Language*, Vol. 59, No. 3, 502-513.

HANKAMER, JORGE. 2012. Ad-Phrasal Affixes and Suspended Affixation. *Cornell Workshop on Suspended Affixation*.

DEAL, AMY ROSE. 2015. Interaction and Satisfaction in phi-agreement. In *Proceedings of NELS 45*, 179-192.

WATANABE, AKIRA. 2006. Functional Projections of Nominals in Japanese: Syntax of Classifiers. In *Natural Language & Linguistic Theory, vol. 24*, 241-306.

KRISHNAMURTI, BHADRIRAJU., GWYNN, JOHN PETER LUCIUS. 1985. A Grammar of Modern Telugu. Delhi: Oxford University Press.

MATTHEWSON, LISA. 2001. Quantification and the Nature of Crosslinguistic Variation. In *Natural Language Semantics, vol. 9*, 145-189.

GREENBERG, JOSEPH. 1977. Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. In *Linguistics at the crossroads*, ed. Adam Makkai, Valerie Becker Makkai, and Luigi Heilmann, 276–300. Lake Bluff, Ill: Jupiter Press.

DOETJES, JENNY. 2007. Adverbs and quantification. *Lingua*, 117:685–720.

JENKS, PETER. 2017. Numeral classifiers compete with number marking: Evidence from Dafing. Handout from LSA 2017.

WHEELOCK, FREDERIC MELVIN. 2011. Wheelock's Latin, 7th edition.

BAKER, MARK. 2008. The Syntax of Agreement and Concord. Cambridge University Press. Cambridge Studies in Linguistics, vol. 115.

PELLETIER, ROSEANNE. 1993. Predication Within Telugu Nominals. In *Grammatical Relations: Theoretial Approaches to Empirical Questions*, pp. 321-340.

BEJAR, SUSANA., REZAC, MILAN. 2009. Cyclic Agree. In *Linguistic Inquiry*, Vol. 40, No. 1, pp. 35-73.

CAMPBELL, AMY. 2012. The Morphosyntax of Discontinuous Exponence (Doctoral Dissertation, UC Berkeley).

To Speak or Not to Speak: A Call for Meaningful Distance-learning Materials for Peripheral Heritage Language Learners

Bri Alexander
University of Arizona

1.0 Introduction

Hundreds of years of trauma and forced assimilation post-European contact have had and continue to have a deep impact on many aspects of Native American life. Displacement, one aspect of the trauma, has severed individuals from their Native communities, causing identity struggles and language attrition as community members lose geographical access to their community and tongue (Crystal, 2000; Harrison, 2007; McCarty & Zepeda, 2010). Despite distance, however, individuals can still identify as a member of their heritage community and, further, attempt to reclaim and reconnect with their heritage community. For Native-identifying individuals without immediate access to cultural activities and other tokens of identity, it is not uncommon for them to become heritage language learners (Hale & Hinton, 2001; Nicholas, 2009; Peter, 2014; Kickham, 2015; White, 2015). As language serves as a bridge between culture and individuals, its usage signifies membership of a person in a specific community (Hymes, 1964; Giles & Johnson, 1987).

Interestingly enough, though, researchers and leaders in language revitalization, reclamation, documentation, and maintenance focus heavily on in-person methods (such as immersion, master-apprentice programs, and in-person classes), effectively leaving out of the narrative methods for distance (or peripheral) learners. With such a large population of displaced Native-identifying individuals lacking access to in-person resources, however, these commonly-studied and –implemented methods do not apply to all or even the majority of Indigenous heritage language learners. Therefore, this study, using an auto-ethnographic account on being a peripheral Cherokee heritage language learner coupled with personal communication from other peripheral Cherokee heritage language learners, investigates how Cherokee-identifying individuals view linguistic resources' successes in their language learning journey, as well as what motivations they have when turning to linguistic resources. This paper first gives context on Cherokee and available resources, then provides background on Indigenous heritage language learners and current methods in language revitalization, followed by methods and discussion on linguistic resources and learner motivations, before finishing with implications for theory.

2.0 Context

Before discussing Cherokee language learners and motivations, it is imperative to first understand the linguistic situation of Cherokee. This section gives a brief overview of the vitality of Cherokee as a whole (as this study focuses on the Cherokee language as a unit, though there are two main dialects: eastern and western), as well as revitalization efforts currently underway.

2.1 *Cherokee language and vitality*

The Cherokee language is an endangered language spoken by approximately 10,000 individuals, largely in eastern Oklahoma, but also in western North Carolina (Cherokee, 2004; Austin, 2008). The sole Southern Iroquoian branch, Cherokee is considered a stand-alone language (Austin, 2008), although it is related to nine Northern Iroquoian languages: Oneida, Onondaga, Seneca, Mohawk, Cayuga, Nottoway, Huron, Susquehannock, and Tuscarora (Hoffman, 1959). While enrollment in all three Cherokee bands totals over 300,000 citizens, less than half live on or near tribal lands (Cherokee, 2004), and with only three percent of the tribe claiming fluency, the language has reached the "definitely endangered" level as

Proceedings of *WECOL 2017*

classified by UNESCO in Moseley (2010). Lewis, Simons and Fennig (2017) rate Cherokee as a 6b on their endangered scale, meaning that although the language is spoken intergenerationally, the language is losing speakers. However, Lewis, Simons, and Fennig's (2017) rating does not take into account actual, recent work done in the community to assess the language situation. Scancarelli (1986) and Montgomery-Anderson (2015), both active in Cherokee language work, have found that there are consistently no speakers under the age of 40 and that the language is not being transmitted in the home. Therefore, a better place to rate Cherokee would be at a 7 on the GIDS Typology (Fishman, 1991), placing Cherokee as a more endangered language than what Lewis, Simons, and Fennig (2017) assert.

2.2 *Current revitalization efforts*

The rate of language attrition might seem shocking to some at first due to the immense variety of resources available to language learners, from online and technological resources to plenty of printed materials and in-person classes. The Cherokee Nation offers online language classes where students can see a live video of a fluent speaker going through a PowerPoint presentation of ten phrases, as well as numerous downloadable resources and an easy-to-use word list (although it is not extensive). The United Keetowah Band of Cherokees has documents with traditional stories available for free. A couple of mobile apps exist, as well as a Cherokee keyboard for smart phones and Microsoft software. Another online class, created by speakers, teaches grammar and tone through uploaded videos and a textbook. For printed materials, a reference grammar has been published, as well as a Cherokee-English dictionary. The Cherokee Phoenix is still circulating, as well as a magazine called *Anadisgoi*, with both featuring the Cherokee syllabary in addition to English. In-person classes range from immersion schools (offered by Cherokee Nation and the United Keetowah Band) to college level courses at Northeastern in Tahlequah, Oklahoma, to immersion camps hosted multi-annually and other community-level classes. Recently, the Cherokee Nation has also developed a master-apprenticeship program, although spaces are extremely limited and competitive to get into.

3.0 Background

In reality, this research began not from a question but rather an observance, that "heritage language learners of Indigenous languages who do not have access to a community of practice will need different linguistic resources." To adequately address this observation, however, it is first necessary to understand what the best practices and common methodologies currently are for learning Indigenous languages, as well as who heritage language learners are and how these learners can achieve their learning goals.

3.1 *Indigenous heritage language learners*

It should come as no surprise that Indigenous heritage language learners come in all varieties, from younger to older, traditional to non-traditional, and in-community to peripheral, just to name a few. As displacement is a primary factor in language attrition (Crystal, 2000; Harrison, 2007; McCarty & Zepeda, 2010), it is probable that most Indigenous communities of practice have individuals who live away from the geographical area of practice but still identify as a community member. Unfortunately for these peripheral heritage language learners, though, they have grown up in a dominant culture which has excluded them from their heritage and, as LaFramboise et al. (1993) asserts, has required Native-identifying individuals to pass as "culturally competent" (p. 396) by adhering to all socially sanctioned behaviors, having a strong identity based in the dominant culture, knowing of all beliefs and values of the dominant culture, using the approved language of the culture, and actively being a full member of the community. Essentially, the dominant culture not only removes Native-identifying individuals from their respective communities, but it further bars them from any identification factors that they could turn to in order to project their heritage identity. Many who still harbor a connection to their Native heritage, however, turn to language learning in adulthood as a gateway to the community and to unlocking their self-ascribed identity (Hale & Hinton, 2001; Nicholas, 2009; Peter, 2014; Kickham, 2015; White, 2015).

Therefore, Native-identifying individuals turning to language learning are, in fact, desiring cultural and communicative competency (i.e., the ability to produce language in a meaningful context) instead of linguistic competency (i.e, the ability to generate a grammatically correct phrase or sentence, regardless of context or meaning). For Indigenous heritage language learners, the importance of language learning rests

not on the ability to formulate grammatical sentences regardless of context, but on being able to have a meaningful conversation, to respond to unique cultural variables, and to be a successful member of society by adhering to the parameters of inclusion imposed by language. With this motive, distance resources can (and often do) lose their function, since most erase the culture and community – the two connections that Indigenous heritage language learners seek most.

3.2 *Common methods for learning Indigenous languages*

As endangered languages gain more spotlight in linguistics and anthropology, research on language revitalization, reclamation, documentation, and maintenance (and, specifically, on best practices for teaching and learning Indigenous languages) is becoming more prominent. Various sources agree that methods that immerse learners solely in the target language are the most effective; immersion-style schools and language nests continually boast reversal of language loss, first in Hawaii and New Zealand, and now in North America (Hale & Hinton, 2001). Additionally, master-apprentice programs are being implemented in multiple Indigenous communities across the globe to much success (Mentor-Apprentice Program, 2008; Olawsky, 2013; National Breath of Life Archival Institute for Indigenous Languages, 2016). Furthermore, language classes put on by the community or by educational institutions (such as universities and high schools) attract heritage language learners and help produce speakers. These three methods flourish for numerous reasons, but most notably, they are all in-person methods where the learner has direct, physical access to a community of practice and/or a language teacher. As of now, research in language revitalization has yet to delve into methods for distance learners or the use of technology in transmitting Indigenous heritage languages.

Although in-person attendance in immersion-style programs is the ideal learning environment for Indigenous languages, it requires learners to have physical proximity to a community of practice (Hale & Hinton, 2001; Mentor-Apprentice Program, 2008; Olawsky, 2013; National Breath of Life Archival Institute for Indigenous Languages, 2016). As distance learners, displaced community members naturally do not have the option to take in-person classes, but must instead rely on distance materials, such as books, internet downloads, or online classes. However, the majority of these resources fail to incorporate interaction between members of a community of practice. Distance heritage language learners, then, have little to no resources that will satisfy their motives for turning to language learning; since they cannot attend in-person programs and most distance materials lack a connection to a community, distance heritage language learners cannot gain cultural or communicative competency like they desire.

4.0 Methods

Ethnographic accounts offer rich glimpses into life, especially into worlds gone unnoticed. Specifically for Native scholars in linguistic anthropology, presenting one's experience and research via auto-ethnography is becoming increasingly more accepted, a shift that gives credence to perhaps otherwise unspoken views. Thus, as a Native-identifying individual who is a member of this study's target demographic, I derive data from my auto-ethnographic account of growing up as a peripheral heritage language learner of Cherokee. In an attempt to include more voices, this research also includes personal communication from two older males and one older female living in geographically distant places. All four of the views presented in this paper are from peripheral heritage language learners of Cherokee who have interacted with a significant amount of the available in-person and distance linguistic resources.

5.0 Data and discussion: Resources

To begin, this paper will discuss which elements of available linguistic resources peripheral heritage language learners of Cherokee view as most successful in meeting their goals. Presented accounts identify having access to a Cherokee community of practice and achieving cultural and communicative competency as most important, even more so than linguistic competency.

5.1 *My account*

My language learning journey has taken me through all manner of resources, from online material to printed sources like dictionaries, story books, and a reference grammar. Thinking at first that I would learn

Cherokee how I have learned other languages, I first turned to using mobile applications and books, which worked well for about a month, but I quickly found myself deeply frustrated with my inability to parse out vocabulary, and I began to yearn for a more profound experience with the language (though I could not identify yet what this was or would look like). I then discovered online classes provided by the Cherokee Nation, and, excited with the prospect of interacting with a speaker and other beginning learners in real-time, I signed up. Due to various reasons, however, the classes did no more for my Cherokee than my previous independent studying. I could not internalize anything greater than *osiyo* ('hello') and, though the teacher was brilliant, I could not connect what he was saying to the language. Not giving up, though, I flew to the Cherokee Nation in Tahlequah, Oklahoma, to visit the teacher who leads the classes. After connecting with the teacher in-person and establishing myself as a Cherokee heritage language learner, the online classes held new meaning for me; instead of being a list of foreign vocabulary, the context behind the language came alive, and I finally began internalizing the language. Linguistic competency was not what I was seeking, and, truthfully, not even possible for me until I demonstrated that I was a community member and developed communicative and cultural competency. The most important aspect of linguistic resources for me, then, has proven to be access to a community of practice and an understanding of the culture behind the language.

5.2 *T's account*

T is an older male who has interacted with numerous resources, but prefers to take online classes – one being the online class that Cherokee Nation offers with a speaker in real-time, and the other being an independent study class that uses a grammar textbook (with the description that it is not to be used to learn conversational Cherokee). T named the Cherokee Nation class with a speaker as a top resource because, "It was great for connecting to a community of learners." Furthermore, T identified a major flaw with the grammar class that was self-paced, saying that there was "no interaction with live humans."

These online resources are arguably the best from a linguistic and student's standpoint. Offering a live fluent speaker, Cherokee Nation's online class imitates an in-person language class (though students can only ask questions and interact with other students via sending a chat message to the class). Additionally, the grammar intensive online class offers rich linguistic knowledge of Cherokee. Combined, both resources should teach students conversational phrases and how to build their own sentences. However, notice that T does not point to the best part of Cherokee Nation's class as having a live speaker to go over phrases or to answer grammatical questions, or even any reason relating to fluency or linguistic elements. Instead, he focuses on how the class creates a bridge to other learners. He further clarifies a drawback in the other class, stating that it has no interaction with live humans. Although the grammar-intensive class can build linguistic competency, it cannot provide community access. His primary interest in classes entirely circumvents linguistic competency and hones in on access to a community of practice.

5.3 *S's account*

S is an older female who primarily uses CDs and Cherokee Nation's online classes to learn Cherokee. She said, "If you can find a person to communicate on a regular basis, it is even more helpful. I am in an area where it is hard to find a person I can just talk to and learn from." Learners use classes and CDs to guide their pronunciation in order to speak Cherokee. However, even when using resources that promote pronunciation and conversations, there is still a gap in resource utility. With the resources that S uses, she can become linguistically competent by learning phrases through the online classes and practicing "correct" pronunciation with the CDs. However, for her, regular communication with a speaker at any level (notice she does not specify a fluent speaker or other learner) is what is most important. She also specifically identifies herself as a peripheral language learner with no immediate access to a community of practice, acknowledging that as a learner, she is at a disadvantage from others with easier access to speakers. In S's case, it is apparent that even when resources can be used for vocabulary learning and pronunciation, the distance between learner and community of practice is still a tremendous issue, and the importance of communicative competence supersedes that of linguistic competence.

5.4 *J's account*

Lastly, J is an older male who has spent many years learning Cherokee by using every resource imaginable. A peripheral language learner who has worked tirelessly to achieve linguistic competency, J is

much further along in fluency than most peripheral learners, including those in this study. Although he cannot specify his most favored resource, he does pinpoint the most important factor for the resource to be successful for him. He stated, "I try to interact with fluent speakers when possible." For J, no resource is greater than the other; using reference grammars and online classes (among others), he does not assign more resource utility to one over the other. Instead, he acknowledges that connecting with fluent speakers is what makes the difference in his language learning journey. Having access to a community of practice outweighs simply having the ability to speak. Again, it is clear that resources offering communicative competency are more desirable than those promoting only linguistic competency.

5.5 *Summary*

All four peripheral heritage language learners of Cherokee present similar views that cultural and communicative competency, as well as access to a community of practice, are more important than achieving linguistic competency. Despite each learner having a unique background with varying demographic information and levels of interaction with available resources, a consensus is still reached that the most important and successful aspect of linguistic resources, regardless of its ability to advance their language proficiency, is the resource offering a connection to real people who are in the same community of practice in which the learner would like to join.

6.0 Data and discussion: Motivations

Closely related to the last section, this discussion focuses on the motivations behind why each of the peripheral heritage language learners is learning Cherokee. According to the presented accounts, the Cherokee language is a direct pathway to the Cherokee culture. Performing language allows individuals to have access to a community of practice and indicates community membership, regardless of geographical distance.

6.1 *My account*

As mentioned in the previous section, when I began learning the Cherokee language, I initially thought my ultimate goal was to become a fluent speaker. If this were the case, though, I could have easily used the resources available; much exists to boost fluency beyond even the novice level, and achieving linguistic competency is still possible regardless of if the resources are decontextualized or not. However, after meeting my language teacher in-person, I realized that I had turned to performance of the Cherokee language – not being a fluent speaker – as a marker of my identity and inclusion in the Cherokee tribe.

Much like my tribal enrollment card which can be produced on the spot or tucked away, my linguistic competency behaved in the same manner: I could speak at people, but I could not use it to prove that I lived in the Cherokee way or viewed the world as Cherokees do. My purpose in learning Cherokee was to connect with my heritage and to gain community inclusion; without a community of practice, though, linguistic competency can only get a learner so far. Neither my self-ascribed nor ascribed identity discomfort dissipated due to my advancements in fluency. It was only upon performing the language in context, interacting with other Cherokees, and gaining access to a community of practice that I understood and internalized the language and ameliorated my identity discomfort. Simply put, my motivation for learning was and is to achieve cultural and communicative competency.

6.2 *T's account*

T shared similar sentiments, saying, "Being Cherokee and speaking Cherokee are highly interconnected to me. A lot of the culture is carried on the language. Knowing the language helps me understand thinking Cherokee." T reveals much about his language ideologies, stating that for him learning Cherokee is virtually synonymous with being a Cherokee cultural participant. Speaking Cherokee aids in his quest to become Cherokee, both in practice and in thinking. T views the Cherokee language as not just a string of words, but rather as a signal of his identity as Cherokee. He does not separate the language from the Cherokee community, but rather performs the language in order to perform culture, signifying that his motivation to learn Cherokee is to become culturally competent.

6.3 *S's account*

Continuing, S said, "My grandmother was Cherokee. When I found out that we were finally allowed to learn about our Cherokee side I reached the daylights out of [the language]. Each time I learned a new language, I made more friendships which could not have been otherwise." Just like myself and T, S equates her language learning to a connection to community and identity. As someone who had no access to her Cherokee side until later in life, she seized the opportunity to find her Cherokee identity specifically by turning to language learning. She asserts that language singlehandedly establishes new interpersonal connections. Regardless of fluency, Cherokee is more than a decontextualized language, but rather a tool to fuse individuals into a community of practice. For S, linguistic competency is not driving her to learn Cherokee – achieving cultural and communicative competency fuels her learning.

6.4 *J's account*

J agrees, stating, "I am motivated to be able to visit churches, both here in North Carolina, as well as in Oklahoma, and teach doctrinally, both from the English scriptures, plus the Cherokee scriptures." Singling out a defined community of practice, J provides an example of geographical and ideological communities of practice in which he wishes to gain membership by performing the Cherokee language. Widely considered to be hotspots of "authentic" Cherokee language and individuals, churches in North Carolina and Oklahoma provide an excellent platform in which to practice one's Cherokeeness in a palpable community. Thus, J performs the Cherokee language to heighten his social standing as an "authentic" Cherokee in this community of practice, and even when linguistic competency has been achieved, J requires a cultural context in which to project his identity. His motivation to learn Cherokee rests within his ability to use the language as an included member in a Cherokee community of practice.

6.5 *Summary*

Again, despite responses from four drastically different learners, the data support a common view that motivation to learn Cherokee for these peripheral heritage language learners is to be able to use and perform language as an indicator of being a member of a Cherokee community. Language brings about an instantaneous connection to the community, regardless of the presence or absence of other identity tokens and the individual's level of fluency. Being Cherokee and speaking Cherokee are viewed as intertwined, so when a peripheral community member attempts to establish themselves as a Cherokee, they are motivated to turn to language learning to provide them the social status they desire.

7.0 Theoretical implications

Current research in language revitalization, reclamation, maintenance, and documentation largely leaves out distance learners from the narrative. However, displacement has separated innumerable language learners from their respective communities of practice, contributing greatly to language attrition. Sometimes compromising more than half of a Native community (as is the case with the Cherokee tribe), peripheral Native-identifying individuals are turning to language learning now more than ever in order to gain community access, and this could cause a shift in language revitalization, reclamation, documentation and maintenance theory and practice by providing brand new insights on learning trends. Additionally, with such a vast demographic needing resources, linguists, anthropologists, and other language workers must heed Shaul's (2014) advice and adapt resources to fill the needs of the community as a whole by creating resources with distance heritage language learners in mind. Endangered languages could see a change in the diversity of resources available if researchers adapt to include peripheral Indigenous heritage language learners.

8.0 Conclusion

Displacement of Native communities has wrought terrible and wide-spread consequences unto the communities, from individual and communal identity struggles to extreme language attrition. Fortunately, however, language actively connects learners to a community of practice regardless of geography. Indigenous heritage language learners, specifically those learning at a distance, use language learning not for building linguistic fluency, but to connect to their heritage culture and gain access to their respective community. Despite the immense displacement of Native-identifying individuals from their heritage

communities who are potential language learners, though, current methodologies in language revitalization hone in on methods that are only practical for individuals who have immediate, geographical access to a community of practice. Furthermore, linguistic resources created with distance language learners in mind disregard learner motivations (e.g., to connect to a community of practice) and focus instead on offering resources that are disconnected from the culture that work to solely boost fluency. Even for Cherokee, which offers resources for peripheral language learners, available resources intend to build linguistic competency but not cultural or communicative competency. It is now evident, though, that, "Heritage language learners of Indigenous languages who do not have access to a community of practice will need different linguistic resources." Therefore, it is strongly suggested that those in the field who are working on language revitalization, reclamation, documentation, and maintenance remember peripheral heritage language learners and build resources accordingly.

References

Austin, Peter. (Ed.) (2008). *One thousand languages: Living, endangered, and lost.* Berkeley: University of California Press.

Cherokee. (2004). In A. Dalby, *Dictionary oflanguages*. London, United Kingdom: A&C Black.

Crystal, David. (2000). *Language death*. Cambridge, UK: Cambridge University Press. 68-90.

Fishman, Joshua. (1991). *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Bristol, PA: Multilingual Matters Ltd.

Garrett, Michael Tlanusta. (1996). "Two people": An American Indian narrative of bicultural identity. *Journal of American Indian Education, 36*, 1-21.

Garroutte, Eva. (2003). *Real Indians: Identity and the survival of Native America*. Berkeley: University of California Press.

Giles, Howard, & Johnson, Patricia. (1987). Ethnolinguistic identity theory: A social psychological approach to language maintenance. *International Journal of Sociology of Language, 68*, 69-99.

Hale, Kenneth, & Hinton, Leanne. (2001). *The green book of language revitalization in practice*. San Diego, CA: Brill.

Harrison, K. David. (2007). *When languages die: the extinction of the world's languages and the erosion of human knowledge*. Oxford, UK: Oxford University Press.

Hoffman, Bernard G. (1959). Iroquois linguistic classification from historic materials. *Ethnohistory, 6*, 160-185.

Hymes, Dell. (1964). Introduction: Toward ethnographies of communication. *American Anthropologist, 66*, 1-34.

Kickham, Elizabeth. (2015). *Purism, prescriptivism, and privilege: Choctaw language ideologies and their impact on teaching and learning* (Doctoral dissertation). Retrieved from OU Dissertations.

LaFramboise, Teresa, Coleman, Hardin, & Gerton, Jennifer. (1993). Psychological impact of biculturalism: Evidence and theory. *Psychological Bulletin, 114(3)*, 395-412.

Lewis, Paul M., Simons, Gary F., & Fennig, Charles D. (Eds.) (2017). *Ethnologue: Languages of the world.* Dallas, TX: SIL International. Online version: http://www.ethnologue.com

McCarty, Teresa, & Zepeda, Ofelia. (2010). Native Americans. In J.A. Fishman & O. Garcia (Eds.), *Handbook of language and ethnic identity* (323-339). New York, NY: Oxford University Press.

Mentor-Apprentice Program (2008). *First People's cultural council.* Retrieved from http://www.fpcc.ca/language/Programs/Master-Apprentice.aspx

Montgomery-Anderson, Brad. (2015). *Cherokee reference grammar.* Norman, OK: University of Oklahoma Press.

Moseley, C. (Ed.). (2010). *Atlas of the world's languages in danger* (3rd ed.). Paris: UNESCO Publishing.

National Breath of Life Archival Institute for Indigenous Languages (2016, September 8). *About the workshop.* Retrieve from http://www.nationalbreathoflife.org/about-the-workshop

Nicholas, Sheilah. (2009). "I live Hopi, I just don't speak it" -- The critical intersection of language, culture, and identity in the lives of contemporary Hopi youth. *Journal of Language, Identity, & Education, 8*, 321-334.

Olawsky, Knut J. (2013). The master-apprentice language learning program down under: Experience and adaptation in an Australian context. *Language Documentation & Conservation, 7*, 41-63.

Peter, Lizette. (2014). Language ideologies and Cherokee revitalization: Impracticality, legitimacy, and hope. *Journal of Immersion and Content-Based Language Education, 2*, 96-118.

Scancarelli, Janine. (1986). Pragmatic roles in Cherokee grammar. *Annual Meeting of the Berkeley Linguistics Society, 12*, 224-234.

Shaul, David Leedom. (2014). Language revitalization and revival. In *Linguistic ideologies of Native American language revitalization: Doing the lost language ghost dance* (pp. 31-44). New York, NY: Springer International Publishing.

White, Louellyn. (2015). *Free to be Mohawk: Indigenous education at the Akwesasne Freedom School*. Norman, OK: University of Oklahoma Press.

The Impact of Visual--Spatial Awareness on the L2 Reading Ability of a Bilingual Child

Afnan Alghamdi
California State University, Fresno

1. Introduction

There is an established relationship between visual-spatial awareness and reading fluency in monolingual children. Based on many studies, there is a lot of evidence that shows us how visual-spatial awareness is important in literacy development of monolingual children. Does visual-spatial awareness also relate to literacy development (reading ability) in a bilingual child? In the present study I will examine if visual-spatial attention is consistent with L2 reading fluency of a bilingual child.

It is necessary to conduct this study, since there are no such ones that have been done concerning the effect of visual-spatial awareness on L2 learning. Most of the studies concentrate on the effects that visual-spatial awareness has on monolinguals. The results will be important as they will be used to reform ways of L2 teaching by either incorporating or ignoring the visual-spatial aspects.

1.1 *Hypothesis*

Visual-spatial awareness is consistent with L2 reading fluency of a bilingual child.

2. Literature Review

There are a number of studies that have been conducted to establish the effect of visual-spatial awareness on the cognition and fluency of L2 learners. In a study conducted by Valentina Tobia and Gian Marco in 2014, they investigated the role of linguistic and visual-spatial attentional processes in predicting how fluent children attending preschool are in reading Italian. The researchers established that the students in the first and second grade performed equally in visual, blending, and vocabulary tasks. There were also significant improvements in the performance of the second and third grade students. The results of the research are important to the study because they reveal the effect of visual-spatial awareness on the learning of a foreign language. However, the results do not reveal any significant relationship between gender and grade. The study confirmed that although vocabulary and VSTM were weak, they were important in making predictions and as the learner become more skilled the role of predictors changes. According to Tobia and Marco, RAN and PA form influential predictors in the prediction of fluency of readers in primary school in reading Italian.

Another study by Hoyoen Kim (2004) reveals that all indicators of visual perception skills were found to be significantly positively correlated with each other. According to the data collected, third graders were found superior to second graders in various measures accessed which included eye movement, reversal, rapid naming and reading fluency. The results indicated that there was a high correlation between eye movement variables and rapid naming variables. Kim asserted that phonological awareness is the most important indicator of determining learning difficulties in students. The results displayed by Kim are consistent with the finding in the research done by Valentina Tobia and Gian Marco. Having a visual-spatial awareness helps in the process of promoting literacy in language learning. Students who have the awareness acquire aspects of a language faster than those who do not depend on it. Therefore, incorporation of visual-spatial aspects in L2 learning can be helpful in promoting the acquisition of the second language. The findings are against Kim's hypothesis providing that visual skills, phonological skills, phonological awareness, and rapid naming do not have any significant relationship with reading fluency.

The results displayed by the study conducted by Kim are also supported by Liu, Chen, and Wang in their findings. They examined the relationship existing between reading fluency and visual-spatial attention. The researchers established that visual-spatial attention is a unique predictor of reading fluency (Liu, Cheng, and Wang 2015). The study demonstrated that the visual-spatial attention of student in the third grade in Hong Kong Chinese increased reading accuracy and spelling of words. The evidence of visual-spatial awareness promoting literacy in language learning has also been supported by Kenneth Kavale and Steven Forness' study where they assessed the relationship between auditory and visual perception as well as reading

achievement. The findings establish a strong relationship between auditory and visual perception as well as reading ability. The results form a foundation through which analysis of the current research will be based. They indicate strongly that visual-spatial awareness is a major contributor to language literacy. It improves the understanding of language because the leaners are able to think artistically about a language.

However, there are very few studies on Arabic-learning children's literacy and visual/spatial development. One of the rare studies was done by Haya al Mannai and John Everatt (2005) where they investigated learners' literacy skills as well as their ability to decode letter strings and measures of phonological awareness. The studies revealed that measures of phonological skills were important predictors of variability both in reading and spelling amongst the Arabic speaking children. The results are in line with the other studies that indicate a relationship between visual/spatial awareness and language literacy.

In another study involving Arabic children that was conducted by Samar Zebian (2005), revealed that cultural artifact in writing have a link in numeric and spatial processes. In monolingual Arabic children, spatialized mental number line is oriented from right to left. English monolinguals have a number line that is oriented from left to right. This implies that culture has an impact on spatial awareness. Bilinguals will have the option of choosing between an orientation of left to right or right to left based on the context.

According to Olson and Bialystok (2014), visual and spatial cognition between monolingual and bilingual children differs because of the representation of everyday perception and action. The mental objects are represented in language by speech. Languages play a representation function of people's visual perceptions. A monolingual child has a more developed visual/spatial cognitive function because they use only one language to represent their mental space. This differs with the bilingual children who have two languages of representation hence can have a measure of variability. Representations in one of the languages may not evoke mental objects in the cognition of the child when they are do not fully understand the language. Different languages have alternative forms of representations that make objects appear dissimilar.

Correlation coefficients aggregated through auditory and visual perceptual skills for individual reading skills, homogeneity tests on all aggregations were determined to be homogeneous. The results showed that reading ability was strongly correlated to auditory and visual skills. The strength of the variables indicated that visual or audio discrimination, visual closure, auditory blending and auditory comprehension increased the ability to predict reading fluency. The authors assert that, although auditory and visual perceptual assessments increase the level of predicting reading fluency, they should not be used routinely in a psychoeducational battery. It was concluded that perceptual processes need not be considered as the primary of reading ability due limitations surrounding the magnitude and the nature of the correlation between reading and perceptual skills.

3. The present study

Since many of the studies investigate the visual-spatial attention to the reading ability on monolingual children, in my project, I aimed to provide a new insight to study the relationship between visual-spatial awareness and L2 reading ability in a bilingual child. Specifically, I examined the relationship between visual-spatial awareness and L2 reading skill in a first grade Arabic-English bilingual child.

I used I.Q. test for children in their first grade. The test was designed to measure visual-spatial awareness. The test consists of 208 questions as multiple choice questions to measure visual-spatial attention and the test written by Dr. William A. McConochie, a psychologist who developed The I.Q. test for kids. (McConochie, 2007). The child was asked to answer all the questions on the I.Q. test within 40 minutes. The test measures three different types; logical ability, mathematical ability, and visual-spatial skills. In my project, I focused on the result of the child's visual-spatial awareness to see if it consistent with her reading ability. Furthermore, the child was asked to read 20 passages designed by DIBELS ORF (Dynamic Indicators of Basic Early Literacy Skills, Oral Reading Fluency Test).

3.1 *Methodology*

3.1.1 *Participants* An Arabic bilingual child participated in this study. The child is 6 years old. She is a first-grade student, and her native language is Arabic. She started to learn English at age of 4. She is now studying both Arabic and English at an International school. However, Hermila Braun first grade's teacher also participates in this project to measure the child's reading fluency.

3.1.2 *Data* The data was collected from I.Q. Online Test and DIBELS ORF test. The I.Q. online test examined at home while the DIBELS ORF test examined at school by Hermila Braun, the child's teacher.

3.1.3 *Procedure* All measures were done in a quiet place, and it took three days to be done.

3.2 *Measures*

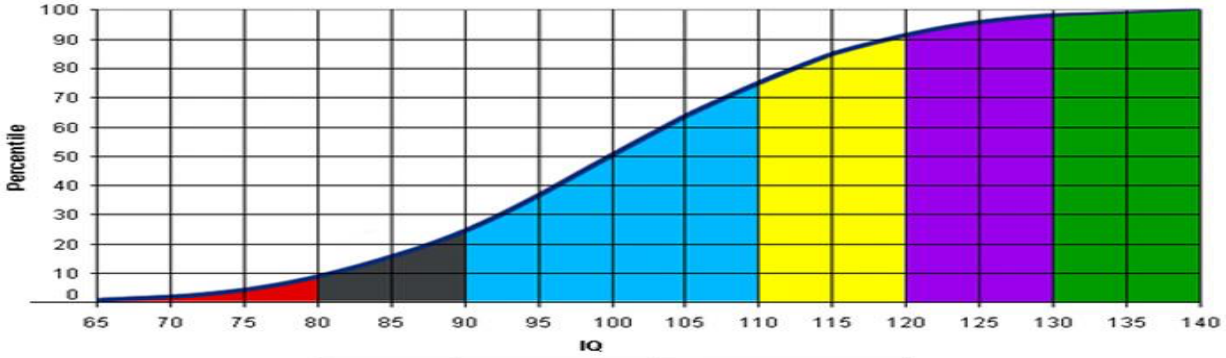
3.2.1 *Visual-spatial awareness* I.Q. Online Test to measure visual-spatial skill for children age of 6 (I.Q. test for kids, McConochie, 2007).

3.2.2 *Reading fluency* The child was asked to read 20 passages within 1 minutes following the all instruction that DIBELS ORF placed to measure the child Oral Reading Fluency. If the child read more than ten words in each passage that qualified her to retell the story within 1 minute.

The purpose for retelling the story is to know if the child has comprehended the reading material. The comprehension ability of the child is not counted in this study. "The result of ORF was measured by dividing the number of words read improperly by the total number of reading words. The ORF rate is the number of correct words per minute" (Good, R.H., & Kaminski, R.A., & Dill, S. 2002, p6).

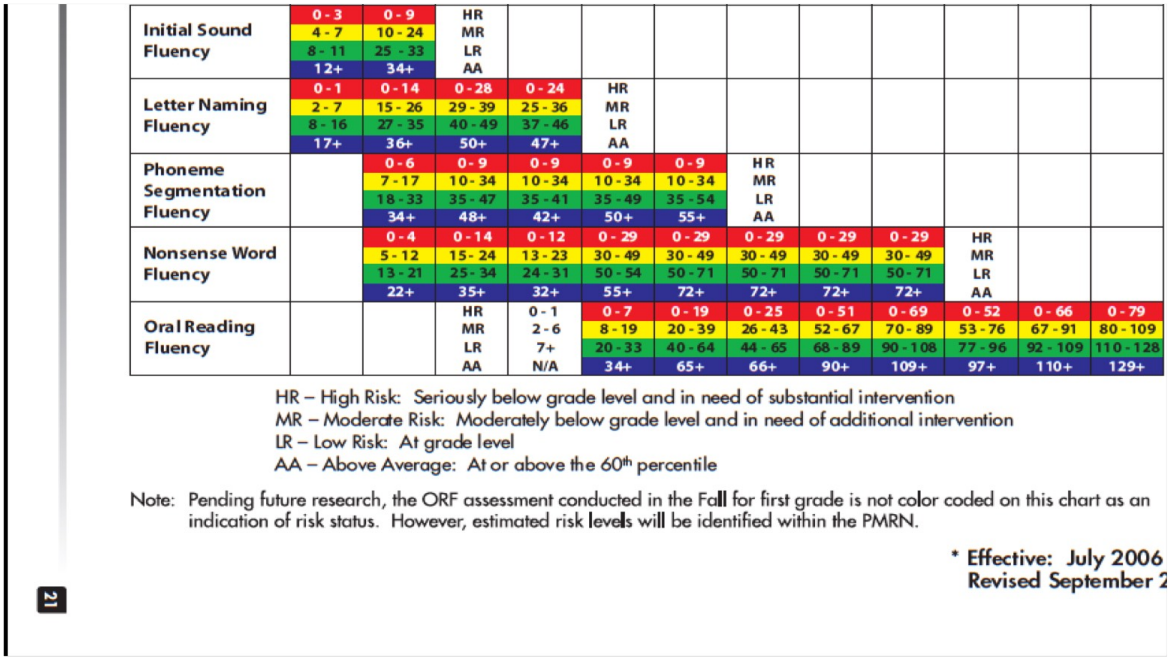
4. Results

The I.Q. test result was 112 scores. The minimum score the student can get to achieve visual-spatial awareness normal average is 90. The result shows that the child's score in visual-spatial awareness is above average. The following graph illustrates the result:



	I.Q.	Intelligence Level
	130 and Above	Very High
	120 to 129	High
	110 to 119	Above Average
	90 to 109	Average
	80 to 89	Below Average
	79 and Below	Low

The result of the child's reading fluency was measured by following the test's measurement of the result. "Once the student read all passages, the median score of the twenty passages was recorded. The minimum score which students can obtain to be considered achieving adequate grade-level reading fluency, is established at 60 correct words per minute" (Good, R.H., & Kaminski, R.A., & Dill, S. 2002, p24). The total percentage was 32,4 (LR= 20-33) which indicates that the child is at the grade level. The following chart illustrates the result:



Initial Sound Fluency	0 - 3	0 - 9	HR									
	4 - 7	10 - 24	MR									
	8 - 11	25 - 33	LR									
	12+	34+	AA									
Letter Naming Fluency	0 - 1	0 - 14	0 - 28	0 - 24	HR							
	2 - 7	15 - 26	29 - 39	25 - 36	MR							
	8 - 16	27 - 35	40 - 49	37 - 46	LR							
	17+	36+	50+	47+	AA							
Phoneme Segmentation Fluency		0 - 6	0 - 9	0 - 9	0 - 9	0 - 9	HR					
		7 - 17	10 - 34	10 - 34	10 - 34	10 - 34	MR					
		18 - 33	35 - 47	35 - 41	35 - 49	35 - 54	LR					
		34+	48+	42+	50+	55+	AA					
Nonsense Word Fluency		0 - 4	0 - 14	0 - 12	0 - 29	0 - 29	0 - 29	0 - 29	0 - 29	HR		
		5 - 12	15 - 24	13 - 23	30 - 49	30 - 49	30 - 49	30 - 49	30 - 49	MR		
		13 - 21	25 - 34	24 - 31	50 - 54	50 - 71	50 - 71	50 - 71	50 - 71	LR		
		22+	35+	32+	55+	72+	72+	72+	72+	AA		
Oral Reading Fluency			HR	0 - 1	0 - 7	0 - 19	0 - 25	0 - 51	0 - 69	0 - 52	0 - 66	0 - 79
			MR	2 - 6	8 - 19	20 - 39	26 - 43	52 - 67	70 - 89	53 - 76	67 - 91	80 - 109
			LR	7+	20 - 33	40 - 64	44 - 65	68 - 89	90 - 108	77 - 96	92 - 109	110 - 128
			AA	N/A	34+	65+	66+	90+	109+	97+	110+	129+

HR – High Risk: Seriously below grade level and in need of substantial intervention
MR – Moderate Risk: Moderately below grade level and in need of additional intervention
LR – Low Risk: At grade level
AA – Above Average: At or above the 60th percentile

Note: Pending future research, the ORF assessment conducted in the Fall for first grade is not color coded on this chart as an indication of risk status. However, estimated risk levels will be identified within the PMRN.

* Effective: July 2006
Revised September 2

21

5. Discussion

The aim of my project was to determine if the established relationship between visual-spatial awareness and reading fluency in monolingual children also holds for a bilingual child. The fact that the child's score on the I.Q. visual-spatial test was above average and the child's score on the DIBELS ORF test was low risk is consistent with the previous findings for monolingual children.
I.Q. above average ≈ reading fluency LR.

6. Conclusion

6.1 *Limitation and implication of my study* Only one bilingual child participated in this study. It seems that if future researches include bilingual children of the same and different age groups, it would be more helpful to understand the relationship between visual-spatial awareness and reading ability in a wide range. Also, the study result would be more powerful if it was a longitudinal one.

References

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Good, R. H., & Kaminski, R. A., & Dill, S. (2007). DIBELS Oral Reading Fluency. In R. H. Good & R. A. Kaminski (Eds.), *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Kavale, K. A., & Forness, S. R. (1996). Social skill deficits and learning disabilities: A meta-analysis. *Journal of Learning Disabilities*, *29*(3), 226-237.

Kim, H. (2004). *The effects of phonological awareness, rapid-naming, and visual skills on early elementary students' reading fluency*. Unpublished doctoral dissertation, University of Florida.

Kavale, K., & Forness, S. (2000). Auditory and visual Perception Processes and Reading Ability: A quantitative Reanalysis and Historical Reinterpretation. *Learning Disability Quarterly*, *23*(4), 253-270.

Liu, D., Chen, X., & Wang, Y. (2016). The impact of visual-spatial attention on reading and spelling in Chinese children. *Reading and Writing*, *29*(7), 1435-1447.

Mannai, H. A., & Everatt, J. (2005). Phonological processing skills as predictors of literacy amongst Arabic speaking Bahraini children. Dyslexia, 11(4), 269-291.

Olson, D. R., & Bialystok, E. (2014). *Spatial cognition: The structure and development of mental representations of spatial relations*. Psychology Press.

Tobia, V., & Marzocchi, G. (2014). Predictors of reading fluency in Italian orthography: Evidence from a cross-sectional study of primary school students. *Child Neuropsychology*, *20*(4), 449-469.

Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing: The SNARC effect and the reverse SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers. *Journal of Cognition and Culture*, *5*(1), 165-190.

Simplification and Complexity in the Interlanguage of Equatorial Guinean Spanish

Omar Beas, Lillie Essah
Arizona State University

1 Introduction

A general observation in language contact situations is that lexical tone is rarely present or functional in creole languages no matter whether the lexifier language is closely related or not (McWhorter 1998). This fact is even replicated in comparable situations such as L2 acquisition in which tonal distinctions in the native languages do not get transferred into the L2 as shown by Musau (1993). The present paper addresses the question of why lexical tones do not get transferred by examining Equatorial Guinean Spanish (henceforth EGS), a bilingual variety (Quilis and Casado Fresnillo 1995, Lipski 2008, Bibang Oyee 2002) deeply influenced by the vernacular tonal languages spoken in the Equatorial Guinea. The following example by Lipski (2008) suggests that EGS distinguishes high (H) and low (L) tones:

(1) el **què tié**ne **dì né**ro **nò há**bla
L H L H L H
"the one who has money does not speak"

yo **pèn sá**ba qu'**ès tá** a **rrí**ba
L H L H H
"I thought that (it) was up(stairs)"

Although not completely functional as in the vernacular languages of Equatorial Guinea, H/L tones in (1) are peculiar because tonal assignment is invariant across the lexicon in EGS, suggesting some degree of phonologization (Lipski 2016).

In order to explain the presence/absence of lexical tone, we resort to the process of tonogenesis in the languages of the world and two correlations shown to influence the absence of tonal distinctions: size of consonant inventory and syllabic structure (Maddieson 2005, 2007. Thus, we hypothesize that lexical tones are not automatically transferred from the vernaculars to EGS, but this progression is an "interrupted process" alike to what is observed in some creoles and language contact varieties (McWhorter 1998, 2007).

The structure of this paper is as follows: Section 1 provides a brief linguistic overview of the languages spoken in Equatorial Guinea in which EGS is described as an interlanguage. In section 2, we elaborate on the presence of lexical tones in EGS and in similar contact situations. Section 3 compares the consonant inventory in Castilian and in the vernacular languages of Equatorial Guinea with EGS, illustrating two processes at work: transfer and substitution leading to a generalized simplification in EGS. Section 4 discusses the structure of the syllable in EGS, which cannot be simply shown to be the result of transfer from the vernacular languages. Finally, section 5 discusses possible implications and directions for further research.

2 Linguistic Overview

Located in the Gulf of Guinea, Equatorial Guinea constitutes the only nation in Africa in which Spanish is spoken as official language. More surprising is, however, to find such diversity of cultures, ethnic groups and languages in a territory with relatively small size of population. According to the National Institute of Statistics in 2015, the population in Equatorial Guinea has been estimated in approximately 1,222,442 inhabitants distributed in the continental and insular areas of the country. In the mainland region, called Rio

Proceedings of *WECOL 2017*

Muni, the majority of Equatorial Guinea's population is concentrated with around 882,747 inhabitants, mostly of Fang origin. In contrast, the insular region is much less populated with 339,395 inhabitants dispersed in several islands such as Annobon, Corisco, Elobey Grande, Elobey Chico and most importantly Bioko. Although the Bubi was the majoritarian ethnic group in Bioko at some point in the history of Equatorial Guinea, this situation has been changing rapidly in the last fifty years due to the increasing political power of the Fang, the dominant group in the country.

2.1 *Vernacular and Official Languages* Beyond these numbers, both the continental and insular regions comprise other smaller ethnic groups, each characterized by its own local language. Thus, along Fang and Bubi, several other languages coexist in the country as illustrated in Figure 1:

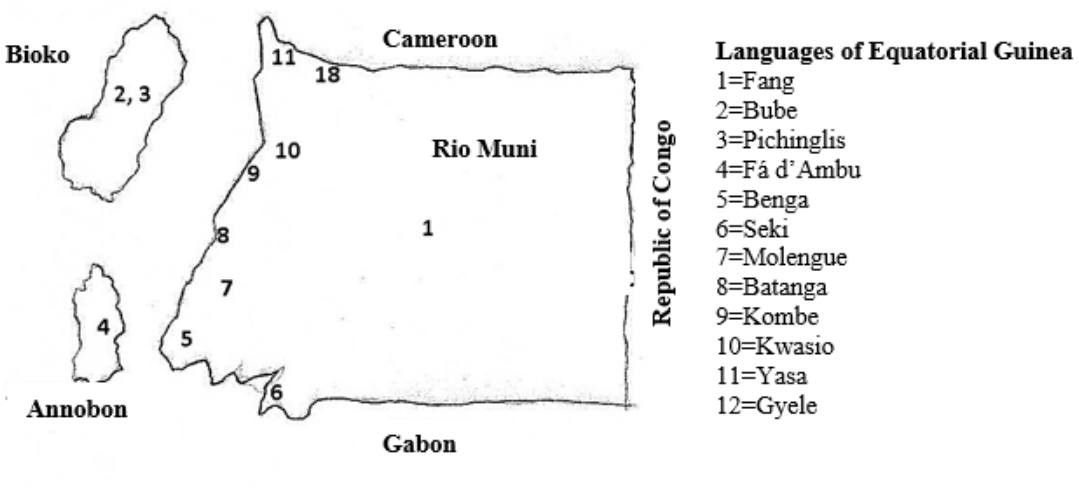


Fig. 1 Languages spoken in Equatorial Guinea

Fá d'Ambu and Pichinglis are two remarkable exemplars of language contact in Equatorial Guinea. Originated from the intermixing of African women slaves and Portuguese, Fá d'Ambu is a Portuguese-lexifier creole spoken in the island of Annobon which remains closely related to the languages spoken in Sao Tome and Angola. In contrast, Pichinglis is an English-based pidgin spoken in the island of Bioko where it is used in all the domains as a vehicle of interethnic communication with the exception of education and government.

The remaining languages in the map are known as the languages of the "playeros", who are geographically relegated to the periphery of Rio Muni. Groups such as the Benga, Seki, Molengue, Batanga, Kombe, Kwasio, Yasa and Gyele were the former inhabitants of Rio Muni but they were displaced to their current location with the expansion of the Fang.

In this scenario, the structure of Equatorial Guinean society closely correlates with the linguistic status of the vernacular languages. Table 1 shows the distribution of the vernacular languages of Equatorial Guinea with respect to the number of L1 speakers and their status as endangered languages as suggested by the EGIDS index (Lewis and Simmons 2010):

Vernacular languages	Number of L1 speakers	EGIDS Scale
Fang	589,000 (72.1%)	3=widely used
Bube	51,000 (6.25%)	6b= at risk
Kwasio	13,000 (1.59%)	6a=active
Seki	11,000 (1.35%)	6a=active
Kombe	9,200 (1.12%)	6a=active
Batanga	9,000 (1.10%)	6b=at risk
Pichinglis	6000 (0.74%)	3=widely used
Fá d'Ambu	5,000 (0.61%)	6a=active
Benga	3900 (0.47%)	6b=at risk
Molengue	1,000 (0.12%)	7=nearly extinct
Yasa	910 (0.11%)	6b=at risk
Gyele	29 (0.003%)	6b= at risk

Table 1. Number of speakers by language and EGIDS index in Equatorial Guinea (Lewis and Simmons 2010)

In a multilingual society such as the one in Equatorial Guinea, the data reported by Table 1 indicates the predominance of Fang among the rest of vernacular languages. Clearly, the sociopolitical expansion by the Fang has created a situation of diglossia in which subordinate ethnic groups are displaced along with their original languages.

2.2 *Spanish in Equatorial Guinea* In addition to the gamut of vernaculars, Indo-European languages are also present in the country. Apart from Fá d'Ambu and Pichinglis, which resulted from the contact of local languages with Portuguese and English, Spanish plays a remarkable role as the national language of Equatorial Guinea spoken by more than 75% of the population (United Nations, Department of Economic and Social Affairs, 2015). Thus, even when Spanish along with Portuguese and French are recognized as official languages by the government,[1] both Portuguese and French are truly foreign languages to most of the population because their speakers are restricted only to a small number of educated people in formal situations.

Curiously, the contact of vernaculars and Spanish in Equatorial Guinea has not evolved into a pidgin or a creole: EGS still remains as a second language amidst several, a very important one. Vernaculars persist in their role as L1s and as a sign of ethnic identity, but Spanish is taught and learned obligatorily when formal education begins in school. In contrast to French and Portuguese, both of limited functional usage, the use of Spanish extends to the political, judicial and administrative domains. With a record 95% literacy rate, according to a recent report from the World Bank, it is not surprising to find around 685,000 speakers of EGS, who are mostly bilinguals with a vernacular language (Ethnologue 2013).

Although EGS neither plays a utilitarian role nor serves as an L1, EGS has been described as a "bilingual dialect with very defined characteristics" (Hualde et al. 2009; Lipski 1999, 2008) or simply as a second language (Vuskovic 2013). Leaving aside this debate, we will refer to EGS using the notion of interlanguage. According to Selinker (1972), an interlanguage results from the process of second language acquisition as a non-linear and fragmentary development, "marked by a mixed landscape of rapid progression in certain areas but slow movement, incubation or even permanent stagnation in others." Modelled after the Castilian variety, EGS is learned during the first years of schooling but not necessarily spoken outside formal contexts or in the context of family interactions, which leads to a greater variability if compared with pidgins or creoles. Figure 2 summarizes the main components in the notion of interlanguage we are adopting in this paper:

[1] Both connections with French and with Portuguese are recent. Whereas French is mostly spoken by Equatoguineans who initially fled to Gabon and Cameroon during the dictatorial rule of Macias Nguema, Portuguese was adopted as official language in 2010 with the purpose of improving the bilateral relations with other African nations in which this language is influential.

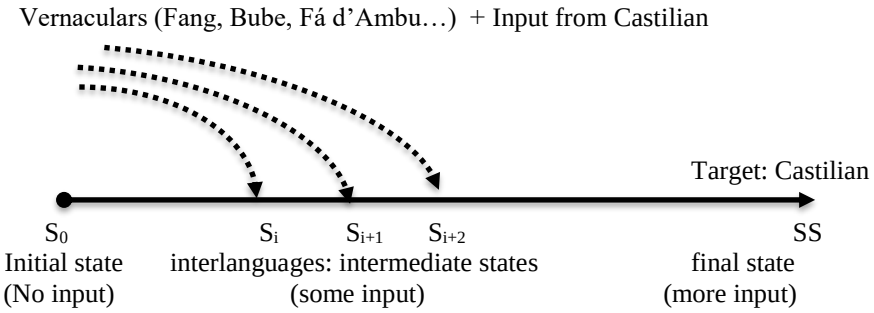


Fig 2. Equatorial Guinean Spanish as a set of interlanguages

In this schematic representation, the concept of interlanguage is characterized by the conjunction of three factors (Major 2001). First, the role of L1 is represented by the vernaculars whose structures may be transferred to EGS. In addition, some processes depicted here as "intermediate states" might not correspond to the L1 vernaculars per se, but instead they could be idiosyncratic or developmental only with respect to the L2. Finally, although still subject to much debate, there are also universal processes that do not correspond to the L1 or to the L2 and in this sense these universals operate as guiding principles even in absence of relevant input. As exposure to Spanish increases, learners move from different levels of interlanguage towards the target or can unexpectedly become stuck or fossilized at some point of development.

3 Lexical Tones in Equatorial Guinean Spanish

In one of the earliest accounts of EGS, Castillo Barril informally identified a "particular intonation" as typical of the Spanish spoken by Equatoguineans (Castillo Barril 1964; 1966: 68). He attributed this "exotic" feature to the influence of the vernacular languages of Equatorial Guinea but without any further precisions (see also Granda 1985, 1986; Granados 1986; Quilis and Casado Fresnillo 1992, for similar observations). For example, the following is reported from a Bube speaker (Lipski 2008):

(2) **Pué**de **dù rár sùs** se**sén tà** años
H L H L H L
"It can last sixty years"
el **ár**bol **nò** tie**nè mà né**ra **dè** de**sà rrò llár**se
H L L L H L L L H
"This tree does not have a way to grow up"
Tènian **tò dà** clas**è** de **à bó**nos
L L L L L H
"(They) had all kinds of fertilizers"

In effect, this intuition about the role of the vernaculars is supported by the two-way distinction of high tones (H) and low tones (L) at the lexical level in Fang, Bube and Fá d'Ambu. In other words, the variations in pitch are meaningful and phonologically distinctive in these languages as shown in (3) reported from a Bube speaker (Bolekia Boleká 1991: 25). Switching from high tones to low tones and viceversa changes the meaning of the lexical item (vowels written with an acute accent (V́) represent high tones, whereas those marked with a grave accent (V̀) represent low tones):

(3)	a.	sààlà	"hunger"	sààlá	"palm peel"
	b.	wèèà	"month"	wèèá	"shrimp"
	c.	ndyòòsì	"log"	ndyòòsí	"snail"

The process by which these peculiar distinctions arise is known as tonogenesis. Across languages, tonal distinctions are derived from the segments surrounding the vocalic nucleus (Hombert et al 1979), namely the structure of the onset or the structure of the coda. Although not the only course of action, the best-known

source involved in the birth of tones is the loss of a voicing contrast in obstruents, particularly in prenuclear position. This means that a simple contrast between H/L tones can be traced to a former voicing distinction in the onset of the syllable. Acoustically, pitch in the syllabic nucleus is naturally lowered with voiced obstruents but raised with voiceless obstruents so in the event of losing this voicing contrast, this subtle pitch difference may be shifted to the vowel.

In the following sections, we will discuss two aspects of the phenomenon in EGS. First, we will identify which specific features are involved in EGS when H/L distinctions occur. Second, if tonogenesis is the linguistic source of pitch distinctions, it is important to understand which factors are correlated with tonal complexity such that they configure the presence or lack of lexical tones in EGS.

3.1 *Transfer and Partial Phonologization* Note that the identification of high tones and low tones in examples such as (1), (2) seems to be fundamentally phonetic and in this sense, the phenomenon is different from what it is observed in vernacular systems such as (3). Nevertheless, even though the phenomenon might not seem systematic, Lipski (2016) argues that partial phonologization must be at stake because the tonal structure in EGS is very stable and with minimal variations across the lexicon.

Furthermore, tonal distinctions with full functionality can be argued to occur in monosyllabic words very much alike to (3). Extrapolating some data from Lipski (2008)'s sample, the following minimal pairs are identified by the H or L tone:

(4) a. nó hablà còn**mí**go — comò estóy ahí me pòngo **mì** pijáma
H — L
"S/he does not speak with me" — "as I am here, I put my pajamas on"

b. Mientràs **él** està aquí èn casa — hàbláran **èl** bubi sólàmèntè
H — L
"While he is in here, at home" — "(they) will only speak (the) Bube language"

Thus, despite the lack of full functionality, EGS shows signs of grammaticalization of certain tonal patterns in specific vocabulary items and a few contrasts in monosyllabic words.

A fundamental question is to identify the structures in which H and L tones are realized in EGS. In this respect, Lipski (2016) hypothesizes that the convergence between Spanish pitch accent and the tonal system of the vernaculars is partially expressed in the identification of the H/L distinctions with the contrast tonic/non-tonic syllables in Castilian. Thus, Equatorial Guinean speakers identify the culminative accent of Castilian with the high tones of their mother tongues, but these speakers would also realize low tones in virtually every non-tonic syllable of their interlanguage.

3.2 *Complexity and Typological Correlations* Let us define grammatical complexity by the metrics in McWhorter (2007), as those linguistic forms that show signs of one or more of the three properties: overspecification (i.e. overt indication of finer semantic distinctions), irregularity (i.e. higher incidence of suppletive forms) or structural elaboration (i.e. higher frequency of occurrence, more rules or number of marked elements).

In phonology, the criterion of structural elaboration is especially relevant as larger phonemic inventories contain more marked, infrequent or rare segments. Similarly, the syntagmatic arrangement of segments into syllables can be as simple as V or CV or as elaborated as CCVCC. Finally, vernacular languages in Equatorial Guinea are more structurally elaborated with respect to lexical tone because their speakers must master finer pitch distinctions, which are absent in other languages.

Simplicity and complexity are two sides of the same coin but with some qualifications. A commonly held assumption is that interlanguage development inevitably leads to structural simplification, at least at initial stages. However, as argued by McWhorter (2007), simplification in one area is many times "counterbalanced by new emergences" in another area. The process of tonogenesis might be understood in this sense. As voicing contrasts in obstruents are inhibited and structural simplification is in order, the translation of this former distinction into pitch contrasts is an instance of complexification.

Curiously, lexical tone is absent in language contact situations and even in second language acquisition.

Lack of distinctive/functional tone is listed as a feature absent or with a low functional load in creoles in McWhorter (1998). Similarly, Musau (1993) has found that lexical tones are never transferred to learners of Swahili as L2 whose L1 is tonal: Nandi, Kamba, Kikuyu, Bukusu, Somali, Maasai or Luo. Despite the presence of partial phonologization and a few minimal pairs, these observations are consistent with what is observed in EGS, which does not show clear indications of using lexical tones in the expected way.

Although few studies have shown which aspects relate to tone complexity, Maddieson (2005, 2007) has discussed how complexity can appear correlated only with some variables but not with others. For example, Maddieson's study found no correlation in the analysis of variance between any of the measures for vowel inventory and syllable complexity but both vowel inventory size and vowel quality size were positively correlated with increasing level of tone complexity ($p<.0001$ F (2, 499)=16.554 and $p<.0001$ F (2, 566)=20.591). Additionally, Maddieson found two other important correlations, which we will take into account in this study in order to explain the emergence or suppression of tone in EGS. First, consonant inventory size was found weakly but positively correlated with tone complexity ($p<.05$ F (2, 566)=3.336), which indicates that whereas relatively smaller consonant inventories are correlated with lower or no tonal complexity, larger inventories tend to appear with some elaboration of tonal distinctions. Second, Maddieson's study has also found a negative correlation between complex syllable structure with lower tonal complexity ($p<.0001$ F (2, 540)=19.15). Thus, although the category of syllable structure does not associate with tonal complexity, if a complex syllable structure with maximal expansion of codas and onsets occurs is likely to occur with lower tonal complexity. We will discuss these two categories in EGS in the following two sections.

4 Simplification, Transfer and Consonant Inventory

In this section, we will compare the salient features of the interlanguage with those of Castilian, Fang, Bube and Fá d'Ambu in order to determine which elements are favored and whether the L1, L2 or the possible presence of universals can account for them. For this purpose, we will be using the descriptive works on Fang by Bibang Oyee (1990) and Alonso Cortes (1991); those on Bube by Levin (1965), Biddulph (1988) and Bolekia Boleká (1991) and the work on Fá d'Ambu by Zamora Segorbe (2010). The data on EGS illustrates what Lipski (2008) would describe as defining features of the Equatoguinean variety, roughly corresponding to an intermediate level of proficiency. Data come from a series of general interest publications by Granda (1985, 1986), Castillo Barril (1964, 1966), Bibang Oyee (2002) and especially the exhaustive work by Quilis and Casado Fresnillo (1992, 1995).

4.1 *Stop consonants* Table 2 shows the distribution of stop consonants in Castilian, Fang, Bube and Fá d'Ambu.[2] Typical phonological inventories across languages frequently display the voiceless stop series /p t k/ including some variability in place of articulation of /t/. Thus, this feature shared among the vernaculars and Castilian is not as surprising as the superficial coincidence in the voiced stop series /b d g/, which we will discuss below. Finally, there are phonological contrasts that are specific only to Bube (palatal segments /c ɟ/) and only to Fang (coarticulated segments /k͡p g͡b/):

	Castilian	Fang	Bube	Fá d'Ambu
Bilabial	p b	p b	p b	p b
Dental	t d		t d	t d
Alveolar		t d		
Palatal			c ɟ	
Velar	k g	k g	k g	k g
Labiovelar		k͡p g͡b		

Table 2. Stops in Castilian, Fang, Bube and Fá d'Ambu

As reported by Quilis and Casado Fresnillo (1992), EGS also displays the contrast between voiceless and

[2] Bibang Oyee (1990) and Bolekia Boleká (1990) mention the existence of geminate consonants in Fang and Bube but given the absence of minimal pairs, we will not consider these segments in our description. See also Alonso Cortes (1991). Similarly, we do not include aspirated stops in Northern Bube reported by Levin (1965) as it does not affect the main conclusion in our paper.

voiced stops but their phonetic realization is also influenced by transfer from the vernaculars, especially Fang. Whereas the phonetic realization of /p t k/ in Castilian is very stable in contrast to the realization of /b d g/, which become fricative in intervocalic contexts, such difference does not arise in EGS. Quite unexpectedly, /p t k/ in EGS show a strong tendency to become voiced in prenuclear position, a process typically found in Fang and Bube (Bibang Oyee 1990, Quilis and Casado Fresnillo 1995) but inexistent in Castilian or in other varieties of Spanish.[3] Moreover, the influence of the vernaculars on EGS also reaches the variable realization found in the articulation of /t/ and /d/ as dental, dentoalveolar and alveolar.

Importantly, transfer from the vernaculars does not apply in every scenario. In particular, the more marked segments in Bube (palatal segments /c ɟ/) and Fang (coarticulated segments /k͡p g͡b/) are never reported in EGS. This gap can be explained as a typological tendency favoring certain classes of segments in the consonant inventory. For example, the occurrence of the set /p t k/ in the phonological systems of the languages of the world suggests that at least one coronal segment must be present (Odden 2014, Hyman 2008). Among coronal consonants (dentals, alveolars, alveopalatals, retroflexes), however, [+anterior] segments are strongly preferred across languages, so this explains why the different realizations of /t/ and /d/ are possible in exclusion of /c/ and /ɟ/ in EGS: /t/ and /d/ are [+anterior] but /c/ and /ɟ/ are not.

4.2 *Affricates and Fricatives* Several contrasts arise when fricatives and affricates are compared in Castilian and in the vernacular languages leading to a quantitative and qualitative difference. To begin, let us examine Table 3 which provides the main segments for each of the four languages. Two preliminary observations are valuable to understand the significance of the data in EGS. First, in quantitative terms, Castilian exhibits a simpler inventory with less number of affricate segments (Castilian=1, Fang=2, Bube=2 and Fá d'Ambu=2) and an almost identical number of fricative segments (Castilian=6, Fang=6, Bube=5 and Fá d'Ambu=7). Second, except for /x/ and /w/, voicing in Castilian is not distinctive in affricate and fricative consonants:[4]

		Castilian	Fang	Bube	Fá d'Ambu
Affricate	Dental		(t͡s) (d͡z)		
	Alveolar		(t͡s) (d͡z)		
	Alveopalatal	t͡ʃ			t͡ʃ d͡ʒ
	Palatal		(t͡s) (d͡z)	t͡s d͡z	
Fricative	Labiodental	f	f v	f	f v
	Interdental	θ			
	Alveolar	s	s z	s	s z
	Alveopalatal				ʃ
	Palatal	ɟ	ɟ	ɟ	ɟ
	Velar	x w	w	x w	w

Table 3. Affricates and fricatives in Castilian, Fang, Bube and Fá d'Ambu

With respect to affricate consonants, Quilis and Casado Fresnillo (1992: 104) reports that the majority of EGS speakers gets very close to the Castilian pronunciation of / t͡ʃ/ but with some variability. Importantly, this approximation to the target language results from the interaction with the vernacular grammars. Bibang Oyee (2003: 15) observes that at initial stages, EGS speakers transfer the voicing distinctions and place of articulation in their vernacular grammars to the interlanguage grammar. Thus, instead of Castilian [ot͡ʃenta]

[3] Perhaps, the only exception to this tendency is found in the EGS spoken in Annobon where Castillo Barril (1966:55) lists a series of phonetic features without mentioning any problems with voiceless stops or stops in general. Worth mentioning is that this perceptual difference may have contributed to the myth that the Spanish spoken in Annobon sounds better than in other parts of Equatorial Guinea, an assertion rejected by Castillo Barril. A better explanation to this perceptual difference should take into account the Portuguese substrate and the degree of isolation of Annobon both from the continent and the rest of islands.

[4] Affricates in the vernacular languages require further remarks. We have found inconsistent descriptions for both Fá d'Ambu and Fang. For example, the pair /t͡ʃ/, /d͡ʒ/ in Fá d'Ambu is also described as palatal in Post et al. (1994) and the pair /t͡s/, /d͡z/ in Fang is reported as dental (Castillo Barril 1965), alveolar (Granda 1986) or palatal (Bibang Oyee 1990, Alonso Cortes 1991). As no consensus has been reached to determine if this is a descriptive problem, or a dialectal variation, we simply note the discrepancies.

"eighty", EGS speakers alternate between [o͡tsenta] and [o͡dzenta] before they can approximate to the pronunciation in the target language.

A similar process is observed for fricative consonants in the interlanguage with a tendency to prefer the simplest grammar. We will concentrate on the case of labiodental fricatives first and then we will discuss the realization of alveolar and interdentals.[5] With respect to the opposition /f/ and /v/, the voicing distinction in the vernaculars is also transferred to EGS, creating a phonological contrast not attested in Castilian or any other Spanish varieties. However, this overdifferentiation does not progress towards the target-like forms as easily as with affricate consonants due to the influence of the written norm in Spanish, which distinguishes the spelling of "v" and "b" even though the language does not contain a /v/ phoneme.

As for the realization of /s/ in EGS, there is some allophonic fluctuation whose origin has not been traced in the vernaculars. To begin with, transfer of the voicing contrast does not seem to occur in EGS but at the same time /s/ admits variations not consistent with the expected Castilian norm. Quilis and Casado Fresnillo (1995) report three realizations of /s/ in the interlanguage of EGS: as predorsoalveolar, as apicoalveolar [s] and as apicodentoalveolar. Interestingly, only the first realization as predorsoalveolar is generalized in the interlanguage instead of the expected apical realization of Castilian. Finally, one of the most difficult segments to master for EGS speakers is the interdental fricative in Castilian /θ/, which is systematically distinguished from /s/: [ˈkasa] "house" versus [ˈkaθa] "hunting", but it is unattested in the vernaculars. Extrapolating from the data by Quilis and Casado Fresnillo (1992), when such distinction occurs in EGS, it corresponds to speakers from urban areas and Rio Muni (i.e. in the continent), suggesting that the effect of formal instruction and level of education are key factors. Nevertheless, many speakers of other areas in the country do not distinguish /s/ from /θ/ and as this segment represents a very rare segment in typological terms, it is not surprising to find that many EGS speakers resort to substitutions. Curiously, the two favorite substitutions for /θ/ are /s/ and /f/, which are typologically unmarked among fricatives.

4.3 *Sonorants* Similar processes of transfer and markedness occur with sonorant consonants in EGS with very interesting precisions. As shown in Table 4, vernacular languages display a richer system with the velar nasal /ŋ/ phoneme in addition to the bilabial, alveolar and alveopalatal segments. In contrast, liquid consonants show the opposite situation: Castilian includes a more elaborated system including the trill consonant /r/ and a palatal lateral /ʎ/; and Fa d'Ambu includes the simplest system with only one liquid consonant:

		Castilian	Fang	Bube	Fá d'Ambu
Nasals	Bilabial	m	m	m	m
	Alveolar	n	n	n	n
	Alveopalatal	ɲ	ɲ	ɲ	ɲ
	Velar		ŋ	ŋ	ŋ
Liquids	Alveolar	ɾ r l	ɾ l	ɾ l	
	Alveopalatal				ɭ
	Palatal	ʎ			

Table 4. Nasals and liquids in Castilian, Fang, Bube and Fá d'Ambu

As for the interlanguage, neither Bibang Oyee (2002) nor Quilis and Casado-Fresnillo (1992) report instances of overdifferentiation of nasals in which the velar nasal /ŋ/ appear in EGS. However, there are some interesting processes that we will detail in Section 4 but it is important to mention that Fang and Bube (but not Castilian) are languages in which nasal consonants can play the role of syllabic nuclei.

The flap /ɾ/ and the alveolar /l/ consonants have been reported to be articulated the same way in the

[5] We will not comment on the palatal fricative /j/ or the velar segments in the paper as EGS speakers seem not to have any problems with these segments. Worth mentioning is that Quilis and Casado Fresnillo (1992, 1995) report different allophonic variations, sometimes influenced by processes existent in the vernaculars.

interlanguage as in Castilian, which makes EGS different from Caribbean Spanish dialects, for example.[6] Something similar can be said with respect to the palatal lateral /ʎ/ but see Quilis and Casado Fresnillo (1992, 1995) for some illustration of depalatalization processes suggesting a preference for [+anterior] segments. A very different situation, however, occurs with the trill consonant /r/, which is hard to attain in EGS not easily solved even at higher levels of education.

5 Developmental factors: The Syllable

Although Maddieson (2005, 2007) did not find a direct association in his sample between tone complexity and syllable structure complexity, he did find that languages with complex syllable structures tend to correlate with lower tonal complexity. In addition, syllable structure complexity is positively correlated with the size of consonant inventory such that the more consonant segments a language displays, the greater it is the increase in complexity in syllable structure. These two findings lead us to expect that if syllable structure shows a "complex" pattern (i.e. one with maximal expansions of onsets and codas), this language is likely to have lower tonal incidence (i.e. a system with no lexical tonal distinctions). We will examine how these two generalizations apply to EGS by comparing Castilian and the vernaculars with the patterns observed in the interlanguage.

In typological terms, Spanish is defined as a complex syllabic language because in addition to the basic CV pattern, this language allows elaborated structures of complex onsets and codas. Table 5 illustrates the main patterns in Castilian:

	Examples	
V	**a**-ro	"ring"
CV	[**mi**]	"my"
CVC(C)	c[**an**]to	"song"
	[i**ns**]tar	"encourage"
CCV	[**gɾa**]no	"grain"
	[**blu**]sa	"blouse"

Table 5. Syllabic patterns in Castilian

Two phenomena should be noted with respect to Castilian syllabic patterns with respect to codas and onsets. First, most examples involving simple or complex codas correspond to very rare cases (Hualde 2005) in which a limited number of segments are allowed. In simple codas, only nasals or liquids are typically allowed (can-to "song", car-ta "letter", al-to "tall") but complex codas are even more restricted allowing only /s/ (ins-ti-tu-to "institute"). Closely related to these restrictions is the lack of contrastiveness in coda position which tolerates neutralization of stop consonants, nasals and rhotics, a process pervasive in all Spanish varieties. Second, complex onsets are restricted in Spanish to the combination of stops and liquids ([kɾe]-ma "cream", [glu]-ten "gluten") and to the combination of voiceless labiodental /f/ and liquids as in [fla]-co "thin", [fɾe]-sa "strawberry").

In contrast to Castilian, syllable patterns in the vernacular languages of Equatorial Guinea are less elaborated. For example, Fang and Bube make regular use of the universally available pattern CV and if simple codas are to be allowed, they are restricted to nasal consonants. No neutralization process operates in coda position in the vernaculars either, which are actually distinctive in stark contrast with Castilian. Finally, the vernaculars can display nasal consonants as syllabic nuclei, an unattested possibility in Castilian. We exemplify the main syllabic patterns of vernaculars using Fang (Alonso Cortés 1991) and Bube (Bolekia Boleka 1991) in Table 6:

[6] The exception is perhaps the EGS spoken in Annobon in which trills or flaps are omitted or substituted by the lateral consonant, presumably due to the influence of Fa d'Ambu.

	Fang examples		Bube examples	
V, C	[**o**]-kí-rí	"tomorrow"	ba-[**u:**]	
			pöö-[**m**]	
CV	[**sá**]	"a tree"	[**mö**]-sondyi:	"lightning"
CVC	o[b**aŋ**]	"a fish"	mo-[**lum**]-bo	
	o[b**am**]	"a hawk"	si-[**pan**]-do	

Table 6. Syllabic patterns in Fang and in Bube

Syllabic structure constitutes one domain in the interlanguage which has been argued to correspond to the operation of linguistic universal. For example, second language acquisition studies such as Broselow, Chen and Wang (1998), Carlisle (2001) and Benson (1988) have shown that there is a strong preference to reanalyze the structures of the target language in terms of CV schemes independently of the structure in the L1. Thus, when the syllabic structure of the L2 is complex (i.e. complex onsets and codas), these studies have shown that learners develop their interlanguage in successive stages closely following the universals proposed by Greenberg (1965):

(5) CV >> CVC / CVCC >> CCVCC
Step 1 Step 2 Step 3

The general pattern shown in (5) also mirrors what was basically found in L1 acquisition (Lleó and Prinz 1996, Levelt, Schiller and Levelt 2000). In both processes of development, the interlanguage first starts from a "universal" CV scheme with a simple onset and without any coda and progressively expands this schema by building complex onsets and eventually simple/complex codas.

Although it is not possible to decide categorically whether there is L1 transfer or development in the syllabic structure of EGS, the data suggests that EGS speakers have moved from Step 1 to Step 2 and are at the beginning of acquiring the complex onsets of Castilian. Thus, the interlanguage shows a progression or development towards the patterns that precisely correlate with lower or no tonal complexity.

Consider first the occurrence of complex onsets in EGS involving liquids. Bibang Oyee (2002) reports that these complex onsets receive a different treatment depending on whether the second consonant is a rhotic or a lateral. Thus, whereas complex onsets involving flap consonants may be simplified by elision, complex onsets with laterals are maintained, suggesting development, not blind transfer from the vernaculars:

(6) a. cu[bɾi]r "to cover" is pronounced as [ku-bí]
b. [tɾa]bajar "to work" is pronounced as [ta-ba-xaɾ]

This intuition is consistent with the development of codas in EGS, which are predicted to appear before than complex onsets but are also predicted to receive a parallel treatment to onset clusters such as those in (6) (Davis and Baertsch 2010). Again, there are two processes in which nasals and liquids are dealt differently when occurring in coda position. If codas are formed by nasals, they are generally retained in EGS as expected if transfer were to operate from the vernaculars. However, codas can alternatively get simplified in order to favor the basic CV schema:

(7) a. al[bun] "album" is pronounced as [al-bum], [al-bũ], [al-bu]
b. co[men] "they eat" is pronounced as [ko-men], [ko-mẽ], [ko-me]

Note that nasalization of the preceding vowel is also an alternative to deal with codas in EGS, even though the rules of nasalization in Spanish only apply in intervocalic contexts (mano [mãno] "hand") or at the beginning of a word followed by a nasal (amo [ãmo] "master"). Once again, the data suggests a conflict between a grammar consistent with the L1 and another in which CV is preferred.

A similar outcome occurs when codas are formed by liquid consonants in EGS. Although these codas are sometimes maintained, simplification towards a more basic CV pattern is also possible and occurs through elision of the liquid consonant (8a)-(8b) or metathesis with the preceding vowel as in (8c)-(8d):

(8) a. [bi̯eɾ]nes "Friday" is pronounced as [bi̯e]nes
b. [aɾ]mario "closet" is pronounced as [a]mario
c. [faɾ]macia "pharmacy" is pronounced as [fa]macia or [fra]macia
d. [feɾ]mentación "fermentation" is pronounced as [fe]mentación, [fre]mentación

Although deserving of a more detailed study, the data suggest that EGS possibly makes use of transfer processes from the vernacular languages, but at the same time EGS goes further by allowing complex onset structures as a last resort and if needed to avoid a more complex structure of codas. In addition, these data also indicate that syllable development is in progress in EGS. Despite the fact that interlanguage structures are more elaborated than those in the L1, EGS speakers are still in the process to acquire the most complex structures of Castilian, which they avoid if possible.

These observations are consistent with Maddieson's generalizations mentioned at the beginning of this section. First, given that complex syllable structures have not fully developed but are in the process to surface in EGS, it is reasonable to expect a lower incidence of lexical tones. Increasing complexity in syllable structure would block a distinctive tonal system and could happen precisely as a function of a relatively larger consonant inventory.

6 Conclusions

In this paper, we have shown that the consonant inventory in EGS is simpler than the corresponding inventories of the vernacular languages and the consonant inventory of Castilian. Unmarked segments such as /p t k/, /b d g/ are retained in EGS but very marked segments are omitted or substituted no matter whether they come from the L1s (labiovelars, palatal stops, velar nasals) or from the L2 (interdental fricatives, trills, palatal liquids). Furthermore, the syllabic structure in EGS displays a process of development towards the structures allowed in Castilian, the target language.

These two findings suggest a preliminary answer to the question of why lexical tones do not emerge in EGS. Given the correlation between consonant inventory size and tone complexity and the correlation between a complex syllable structure and tone complexity (Maddieson 2007), lexical tones do not naturally arise in the interlanguage because the process is "interrupted" by the structures of the interlanguage under development. Both a shorter consonant inventory and a complex syllabic structure conspire against tone elaboration. The lack of distinctive tone emerges in this context as a trait typical of creoles (McWhorter 1998), suppressed by the effects of external factors such as formal instruction.

References

Alonso Cortes, Angel (1991) Fonología segmental del fang ntumu. Filología Románica, 8, 179-212.
Benson, Bronwen. (1988). Universal preference for the open syllable as an independent process in interlanguage phonology. Language Learning, 38(2), 221-235.
Bolekia Boleká, Justo. (1991). Curso de lengua bubi. Malabo: Centro Cultural Hispano-guineano.
Bibang Oyee, Julian. (2002). El español guineano: interferencias, guineanismos. Malabo. Centro Cultural Hispano-guineano.
Bibang Oyee, Julian. (1990). Curso de lengua fang. Malabo. Centro Cultural Hispano-guineano.
Biddulph, Joseph. (1988). Fernandian: the Bubi Bantu language of Bioco/Fernando Po: some notes and commentary. Fernandian Languages Information Centre.
Broselow, Ellen, Su-I Chen and Chilin Wang. (1998). The emergence of the unmarked in second language phonology. Studies in second language acquisition, 20(2), 261-280.
Carlisle, Robert (2001). Syllable structure universals and second language acquisition. International Journal of English Studies, 1(1), 1-19.
Castillo Barril, Manuel. (1966). La influencia de las lenguas nativas en el español de Guinea. Archivo de Estudios Africanos, 20, 46-71.

Castillo Barril, Manuel (1964): "El español en la Guinea Ecuatorial". Español Actual 3.
Davis, S., & Baertsch, K. (2010). On the relationship between codas and onset clusters. In Handbook of the syllable (pp. 71-98). Brill.
Granados, Vicente. (1986). Guinea: del `falar guineu' al español ecuatoguineano. Epos 2.125-137.
Granda, German. (1986). Estudios de lingüística afro-románica (Vol. 1). Universidad de Valladolid, Secretariado de Publicaciones.
Granda, German (1985). Fenómenos de interferencia fonética del Fang sobre el español de Guinea Ecuatorial: consonantismo. Anuario de Lingüística Hispánica 1.95-114.
Greenberg, Joseph. (1965). Some generalizations concerning initial and final consonant sequences. Linguistics, 3(18), 5-34.
Hombert, Jean Marie, John Ohala and William Ewan. (1979). Phonetic explanations for the development of tones. Language, 37-58.
Hualde, Jose Ignacio. Olarrea, Antxon, Escobar, Anna Maria. and Catherine Travis. (2009). Introducción a la lingüística hispánica. Cambridge University Press.
Hualde, Jose Ignacio. (2005). The Sounds of Spanish. Cambridge University Press.
Hyman, Larry. (2008). Universals in phonology. The linguistic review, 25(1-2), 83-137.
Lewis, Paul., and Gary Simons. (2010). Assessing endangerment: Expanding Fishman's GIDS. Revue Roumaine de Linguistique, 55(2), 103-120.
Levelt, Clara, Niels Schiller and Willem Levelt. (2000). The acquisition of syllable types. Language acquisition, 8(3), 237-264.
Levin, Norman. (1965). Notas para un estudio del fonetismo del bubi (dialecto batete). Revista de Filología Española, 48(3/4), 407-413.
Lipski, John. (2016). "Toned-up" Spanish. Romance Linguistics 2013 Romance Languages and Linguistic Theory, 233-256. doi:10.1075/rllt.9.13lip
Lipski, John. (2008). El español de Guinea Ecuatorial en el contexto del español mundial. En: La situación actual del español en África, ed. por Gloria Nistal Rosique, and Guillermo Pié Jahn, 79–117. Madrid: Casa de África/SIAL.
Lipski, John. 1985a. The Spanish of Equatorial Guinea. Tubinga: Max Niemeyer.
Lipski, John 1985b. Contactos hispano-africanos: el español guineano. Anuario de Letras 23.99-130
Lleó, Conxita and Michael Prinz. (1996). Consonant clusters in child phonology and the directionality of syllable structure assignment. Journal of child language, 23(1), 31-56.
Maddieson, I. (2007). Issues of phonological complexity: Statistical analysis of the relationship between syllable structures, segment inventories and tone contrasts. In: Solé, Maria-Josep, Patrice S. Beddor, and Manjari Ohala. Experimental Approaches to Phonology, Oxford University Press.
Maddieson, Ian. (2005). Correlating phonological complexity: data and validation.
McWhorter, John. (2007). Language interrupted: Signs of non-native acquisition in standard language grammars. Oxford University Press.
McWhorter, John. (1998). Identifying the creole prototype: Vindicating a typological class. Language, 788-818.
Musau, P. M. (1993). Aspects of Interphonology: The study of Kenyan learners of Swahili. Bayreuth, Germany: Bayreuth University.
Post, Marike. (1994). Fa d'Ambu. In Arends, J., Muysken, P., and Smith, N. (Eds.). (1994). Pidgins and creoles: An introduction (Vol. 15). John Benjamins Publishing, 191-204.
Quilis, Antonio and Celia Casado-Fresnillo. (1995). La lengua española en Guinea Ecuatorial. Madrid: Universidad Nacional de Educación a Distancia.
Quilis, Antonio and Celia Casado-Fresnillo. (1992). Fonología y fonética de la lengua española hablada en Guinea Ecuatorial. Revue de linguistique romane, 56(221-222), 71-89.
Major, Roy (2001). Foreign accent: The ontogeny and phylogeny of second language phonology. Routledge.
Selinker, Larry (1972) Interlanguage. IRAL, 10, (3), 209-231
Vuskovich, Matthew. (2013). A contrastive analysis of Spanish and Fang: An L2 approach to Equatorial Guinean Spanish as spoken in Bata City. Georgetown University.
Zamora Segorbe, Armando (2010). Gramática descriptiva del fá d'ambô. Madrid: Ceiba Ediciones

Case Stacking as a Post-syntactic Operation

Yitong Chen
Carleton College

1 Introduction

Case stacking refers to the construction where a DP simultaneously displays two (or more) case markers in the same structural position (Levin, 2014). For instance, in the Lardil example (1) below, the nominal *marun* 'boy' is marked with both a genitive case suffix *-ngan* and an instrumental case suffix *-ku*.

(1) *Ngada latha karnjin-i marun-ngan-ku maarn-kur*
I spear wallaby-ACC boy-**GEN**-**INSTR** spear-INSTR
'I speared the wallaby with the boy's spear.' (Richards, 2013, p.43, example (3))

A related construction referred to as case alternation occurs when only one of these case markers is displayed at a time (Levin, 2014)[1]. An example of case alternation is found in Korean. As shown in example (2), the nominal *Jon* can be marked with either a genitive case suffix *-uy* or a nominative case suffix *–i*, but not both at the same time.

(2) *Jon-uy/-i son-i cakta.*
John-**GEN/NOM** hands-NOM small
'John's hands are small/John has small hands.' (Wunderlich, 2014, p.347, example (12a))

As pointed out by Levin (2014), case stacking and alternation pose great challenges to Chomsky's Agree model of case assignment (Chomsky 2000, 2001), according to which each nominal should be uniquely case-marked. The Agree model's account of case stacking and alternation assumes these phenomena to be purely syntactic. In this paper, however, I argue for a *post-syntactic* analysis of both case stacking and case alternation. Under the assumption that case assignment happens within the syntax part of the grammar, I propose that both the case-stacked nominal in (1) and the case-alternating nominal in (2) receive their additional case particles through a process called case concord, akin to how adjectives receive cases from the NPs that they modify. This case concord process occurs post-syntactically in the morphology part of the grammar. Then within phonology, I propose a constraint-based model using the Optimality Theory to account

* An earlier version of part of this paper was presented at the Minnesota Undergraduate Linguistics Symposium in 2017. I thank the participants, especially Dr. Chongwon Park, for their constructive feedbacks and helpful comments. I would like to express my deep gratitude to Dr. Cherlon L. Ussery and Dr. Catherine R. Fortin for all the support and guidance they provided me on this project. I am also grateful to my language consultants who contributed keen insights to my data. Part of this research has been funded by the Carleton Parents' Fund for Academic Excellence.

[1] In this paper, I am adopting a purely morphological definition of case stacking and alternation. In other words, abstract stacking that is not realized morphologically will not be considered as case stacking in the context of this paper.

In addition, the case alternation constructions that are considered in this paper are constructions where the alternation between cases does *not* imply a difference in the structure of the sentence. For example, Lees (2015) examines what he calls "case alternation between accusative and partitive" in modern Finnic languages. For the constructions that he examines, the alternation between accusative and partitive case on the object implies the "boundedness of the clause" (Lees, 2015, p.35). The analysis proposed in this paper does not aim to account for such instances of "case alternation".

Proceedings of *WECOL 2017*

for the cross-linguistic variations regarding case realization[2]. More specifically, whether a language exhibits case stacking and/or alternation is a result of its unique ranking of constraints that determine which case particles are spelled out in the surface structure.

This post-syntactic treatment of case stacking and alternation thus divorces a seemingly purely syntactic process from the (narrow) syntax part of the grammar. Treating case stacking and alternation as results of post-syntactic operations allows us to better capture the different case systems and different patterns of case realization across languages.

In what follows, I will consider the crucial observations on case stacking and alternation constructions that motivate my analysis. In Section 2, I will examine Lardil and Warlpiri, which have been argued to be genuine case stacking languages. While Korean has also been identified as a case-stacking language in previous research, I observe that the Korean case stacking constructions exhibit a different pattern from those in other known case-stacking languages. Therefore, following Schütze (2001), I argue that "case stacking" in Korean is in fact a focus strategy and should be reanalyzed as what I call *pseudo case stacking.* Pseudo case stacking is similar to a case stacking construction on the surface, but only one of the case morphemes attached to the nominal is a genuine case morpheme. Evidence that supports this argument will be presented in Section 3. Section 4 illustrates the mechanism of the case concord analysis of case stacking and alternation. Section 5 introduces the post-syntactic, constraint-based model that aims to capture the cross-linguistic variations of case realization, and illuminates a language typology predicted by this model.

2 Case Stacking

In a typical case stacking construction, both (or all) of the stacked morphemes are identified as case morphemes. Case stacking has been found in a few nominative-accusative languages, including the aforementioned Australian language Lardil. In particular, Lardil allows a possessor nominal to be doubly marked with a genitive case suffix and the case suffix of its possessee (Richards, 2013). This is shown in example (1) (reproduced here as example (3a)) and example (3b). The double case marking in example (3a) signifies the status of the nominal *marun* 'boy' as the possessor of the instrument *maarn* 'spear'. In example (3b), the nominal *marun* 'boy' is the possessor of the direct object *kantha* 'father' and is thus doubly marked with both genitive and accusative case morphemes.

(3) a. *Ngada latha karnjin-i marun-ngan-ku maarn-kur*
I spear wallaby-ACC boy-**GEN**-**INSTR** spear-INSTR
'I speared the wallaby with the boy's spear.' (Richards, 2013, p.43, example (3))
b. *Ngada kurri marun-ngan-i kantha-n*
I see boy-**GEN**-**ACC** father-ACC
'I saw the boy's father.' (Richards, 2013, p.48, example (10a))

In addition to nominative-accusative languages, case stacking is also found in the split-ergative language Warlpiri. In Warlpiri, a locative nominal, which denotes spatial locations, can have an additional ergative or dative case (Simpson, 1991). The nominal *parraja* 'coolamon' in example (4a), for instance, is marked with both the locative case suffix *-rla* and the dative case suffix *-ku*.

(4) a. *Karnta-ngku ka-rla kurdu-ku miyi yi-nyi parraja-rla-ku.*
Woman-ERG PRES-3DAT baby-DAT food.ABS give-NPST coolamon-**LOC**-**DAT**
'The woman is giving food to the baby (who is) in the coolamon.'
b. *Karnta-ngku ka-rla kurdu-ku miyi yi-nyi parraja-rla.*
Woman-ERG PRES-3DAT baby-DAT food.ABS give-NPST coolamon.ABS-LOC
'The woman is giving food (which is) in the coolamon to the baby.'
(Simpson, 1991, p.206, example (187b, c))

[2] Nakamura (1997) also proposes a constraint-based typology of case systems and examines a set of data that overlaps with what is examined in this paper. However, Nakamura (1997) defines case system in terms of the nominals' semantic roles, thematic relations and macroroles, which greatly differs from the definition that I assume in this paper. Our proposed analyses are also quite different despite that both use the Optimality Theory framework.

Compare the examples (4a) and (4b). While (4a) clearly shows case stacking, (4b) is not considered a case stacking construction under our definition. This is because in example (4b), only one morphologically overt case is attached to the nominal *parraja*, even if two cases (absolutive and locative) are assigned in syntax. This difference in case marking results in different interpretations of (4a) and (4b) despite that the two sentences are identical otherwise: in example (4a), the case-stacked nominal *parraja* modifies the indirect object *kurdu* 'boy', while in example (4b), the same nominal modifies the direct object *miyi* 'food'. This suggests that the modification relationship between the case-stacked nominal and the dative-case marked indirect object in example (4a) is in fact what allows the stacked case morpheme to appear.

Critically, as we relate the Warlpiri example (4b) back to the Lardil examples in (3), we note that in all of these case-stacking constructions, the case-stacked nominals are embedded within a DP headed by another nominal. The structural positions of these DPs in which the case-stacked nominals are embedded determine the additional case morphemes that are attached to them.

Yet this observation does not seem to hold for one other language that has been traditionally identified as a case-stacking language, namely Korean. As will be shown in the next section, I argue that the case stacking constructions in Korean is in fact *pseudo case stacking* and thus does not constitute an exception to this observation.

3 Pseudo Case Stacking and Case Alternation in Korean

3.1 *Background* Korean is a nominative-accusative language with overt nominative and accusative case markers[3]. Relevant to our discussion here, Korean also contains a dative case suffix *-eykey*. This dative case morpheme is used to mark the indirect object of a ditransitive verb, or the subject of a psych verb—verbs that describe a psychological experience. Both usages of the dative case are reasonably common across languages. Interestingly, however, Korean seems to also allow a dative-case marked subject to be doubly marked with a nominative case suffix, resulting in a case stacking construction as shown in example (5).

[3] The nominative marker in Korean has three surface forms, *-i, -ka* and *-kkeyse*. *-i* is used following a consonant, and *-ka* is used following a vowel as shown in examples (1) and (2). *-kkeyse* is an honorific nominative marker that is used to mark a nominal representing an esteemed or honored person (Lee&Ramsey, 2000), as shown in example (3).

(1) *Yengi-ka hakkyo-ey kanta*
Yengi-NOM school-LOC enter
'Yengi is going to school.' (Lee&Ramsey, 2000, p.141, example (5a))

(2) *Cangmi-kkoch-i alumtapta.*
rose-flower-NOM pretty
'The roses are beautiful.' (Lee&Ramsey 2000, p.140, example (1))

(3) *Sensayngnim-kkeyse capswusi-ess-ta.*
teacher-HON eat.HON-PAST-DECL
'The teacher ate.' (Sells, 1995, p.293, example (28a))

The accusative marker also has two surface forms, *-ul* and *-lul.* The former is used following a consonant and the latter following a vowel, as shown in examples (4) and (5).

(4) *Swuni-ka chayk-ul ilknunta.*
Swuni-NOM book-ACC read.
'Suni is reading a book.' (Lee&Ramsey, 2000, p.147, example (1a))

(5) *Swuni-ka cha-lul masinta.*
Swuni-NOM tea-ACC drink.
'Suni is drinking tea.' (Lee&Ramsey, 2000, p.147, example (1b))

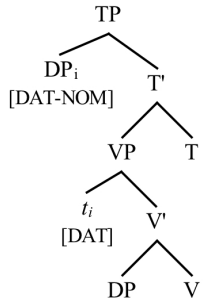
In what follows, I will sometimes refer to the nominative marker as *-ka* and the accusative marker as the *-lul*, ignoring the phonological variations for simplicity.

(5) *Na-eykey-(man)-i paym-i mwusepta*.
I-**DAT**-(only)-**NOM** snake-NOM fearful
'(Only) I am afraid of snakes.' (J. Yoon, 1996, p.110, example 2(a))

An important contrast between the Korean example (5) and the Lardil and Warlpiri case stacking constructions exemplified in (3) and (4) is that the Korean case-stacked nominal *na* 'I' is not embedded in any DP headed by another nominal, nor is it in any modification relationship with the other nominative-case marked nominal *paym* 'snake'.

Previous analyses of Korean case stacking constructions exemplified by sentence (5) divide into two camps—the case marker analysis and the focus marker analysis. The case marker analysis considers both of the stacked morphemes in (5) as case morphemes, just like their counterparts in Lardil and Warlpiri. This analysis was first proposed by Gerdts and Youn (1989), who argue in the Relational Grammar framework that the inherent dative case in example (5) is licensed by the semantic role of the nominal as an experiencer, and the structural nominative case is licensed as the nominal advances to become a subject in the final output of the grammar. Framed within Government and Binding Theory, the case marker analysis essentially argues that the inherent case on the case-stacked nominal is licensed by its deep structure role, and the structural case is licensed by its surface structure role (J. Yoon, 1996). As noted by J. Yoon (1996), non-nominative subjects in Korean, such as the case-stacked dative subject in example (5), have the same properties as the canonical nominative subjects. For instance, non-nominative subjects undergo raising in a passive construction, control honorific agreement on verbs, and bind reflexives (J. Yoon, 1996). Based on these observations, J. Yoon (1996) suggests that while the case-stacked non-nominative subjects are base-generated in a position in VP, they undergo A-movement to the subject position, i.e. the [Spec, TP] position. Movement into the subject position is possible because the psych verb predicates in these case stacking constructions are unaccusative, and thus the [Spec, TP] position will be vacant. Such a movement is shown in (6).

(6)



According to J. Yoon (1996), the double case marking on a Korean nominal is a consequence of this movement, i.e. the dative-case marked nominal advances from it base position to the [Spec, TP] position, where it is assigned the additional nominative case.[4]

Schütze (2001) proposes an alternative analysis of the stacked morpheme in Korean. He argues instead that the stacked morpheme in a Korean case stacking construction is a discourse particle allomorphic to the nominative case marker and is used to mark focus. In other words, Korean case stacking is a focus strategy rather than genuine case stacking.

Following Schütze (2001), I argue that what looks like case stacking in Korean is in fact a focus marker stacked upon a case marker. Therefore, this construction should be re-identified as pseudo case stacking. Below, I review part of Schütze's argument and provide additional evidence to support this analysis.

[4] In a later paper, J. Yoon (2004) proposes that nominative case stacking in Korean is a mechanism to mark the 'Major Subject' status of the non-nominative subjects. He argues that Major Subjects should be differentiated from grammatical subjects, although under some circumstances, a nominative-case stacked Major Subject might be co-indexed with a grammatical subject. Given the limited space and scope of this paper, I will not get into a further discussion on J. Yoon's Major Subject analysis, but the arguments against the case marker analysis discussed below should still stand even if we take the Major Subject analysis into consideration. The reader may refer to J. Yoon (2004) for the details of this approach to account for Korean case stacking constructions.

3.2 *Pseudo case stacking* Schütze (2001) essentially provides two types of evidence to defend the focus marker analysis. First, the stacked morphemes and the unstacked case morphemes behave quite differently in many respects, and thus the stacked morphemes should not be identified as case markers. Second, the stacked morphemes show a similar pattern to the focus/topic marker in Korean; therefore, Korean case stacking should be reanalyzed as a focus strategy. For simplicity, I will focus the subsequent discussion on the stacked *-ka* morpheme, which is allomorphic to the nominative case marker, but the same analysis applies to the stacked *-lul* morpheme and the corresponding accusative case marker. To signal the "case stacking", I will continue to gloss the stacked *-ka* morpheme as NOM even though I argue that it should not be identified as a case morpheme.

First, while the stacked *-ka* morpheme is completely optional, the unstacked case morpheme can only be dropped in very limited occasions such as casual communications (Schütze, 2001). Compare examples (7a) and (7b). As reported in the literature, the native speakers that I consulted indicate that dropping the case markers in (7a) is unnatural but dropping the *-ka* marker in (7b) is allowed in their grammar.

(7) a. *Yengi-??(ka) hakkyo-ey kanta*
Yengi-NOM school-LOC enter
Yengi is going to school. (Lee&Ramsey, 2000, p.141, example (5a))
b. *Na-eykey-(man)-i paym-i mwusepta.* [NOM = FOC]
I-**DAT**-(only)-**NOM** snake-NOM fearful
'(Only) I am afraid of snakes.' (J. Yoon, 1996, p.110, example 2(a))

Schütze (2001) also notes that the stacked *-ka* and the unstacked case morpheme differ in their abilities to control case concord on floating quantifiers. As shown in example (8a), the unstacked nominative subject *haksaygtul* 'student' triggers obligatory case concord on the floating quantifier *seys* 'three'. In example (8b), the case stacked dative subject also controls case concord, consistent with J. Yoon's observation that non-nominative subjects behave like canonical subjects in Korean. Yet the floating quantifier can *only* take the dative case of the subject. Concord with the stacked nominative case morpheme, on the other hand, is ill-formed.

(8) a. *Haksaygtul-i ecey seys-i ttenassta*
Students-NOM yesterday 3-NOM left
'Three students left yesterday.
b. *Haksaygtul-eykey-ka ton-i seys-eykey/*i philyohata* [NOM = FOC]
Students-**DAT**-**NOM** money-NOM 3-DAT/*NOM need
'Three students need money.' (Schütze, 2001, p.201, example (10-11))

While we observe these differences between the behavior of the stacked *-ka* morpheme and the unstacked nominative case marker, there appear to be many similarities between the function and properties of the stacked morpheme and the focus/topic marker in Korean. Korean uses the contrastive suffix *-nun* (or its phonological variation *-un*) to mark contrastive topic or focus, depending on the discourse function of the utterance (Vermeulen, 2012). This is shown in example (9) below. Example (9) can be used either as an answer to the request "Tell me about this hat?" where the *-nun* marker functions as a topic marker, or as an answer to the question "What did John buy yesterday?" where *-nun* functions as a focus marker to signify the new information.

(9) *ku moca-nun$_i$ John-i ecey t_i sasse.*
this hat-NUN John-NOM yesterday bought.
'John bought this hat yesterday.' (Vermeulen, 2012, p.87, example (15a))

Importantly, native speakers that I consulted indicate that the stacked *-ka* morpheme can also induce a focus interpretation on the nominal to which it is attached. The stacked nominal is often the highlighted or new piece of information with respect to the rest of the sentence. Moreover, the stacked morpheme and the *-nun* marker share the same morphological slot and are mutually exclusive (Martin, 1992). For example, my native speaker consultants consider (10a) and (10b) to be two ways to emphasize the nominal *haksayng* 'student' while (10c) is completely illicit.

(10) a. *Haksayng-tul-eykey-ka ton-i philyohata* [NOM = FOC]
Student-pl-**DAT**-**NOM** money-NOM need
'The students need money.' (Gerdts&Youn, 1986, p.16, example (19c))
b. *Haksayng-tul-eykey-un ton-i philyohata.*
Student-pl-DAT-UN money-NOM need
'The students need money.'[5]
c. **Haksayng-tul-i-nun ton-i philyohata.*
Student-pl-NOM-NUN money-NOM need

The above evidence supports the argument that the *-ka* morpheme can have a focus function, yet the question remains, is it possible that the stacked morphemes are case markers with an optional [+focus] feature? While Vermeulen (2012) noted that the Korean non-contrastive foci usually bear case markers, there is some independent evidence suggesting that there should exist focus markers *-ka* and *-lul* that are independent from the structural case morphemes. First, *-ka* and *-lul* can be attached to multiple positions that are rarely structurally case-marked. Schütze (2001) cites an example where *-ka* is attached to a locative-case marked nominal, as shown in (11).

(11) *Ecey-pwuthe-ka nalssi-ka cohaciessta.* [NOM = FOC]
yesterday-LOC-**NOM** weather-NOM good.become
'From yesterday the weather became good.' (J.Yoon, 1987, 156)

The reader might recall that in the Warlpiri case stacking example (4a), a locative-case marked nominal is also doubly marked with ergative case. The crucial difference between Korean and Warlpiri, as previously noted, is that the Warlpiri locative nominal is embedded in a higher level DP as the adjunct of another nominal, and that DP as a whole is assigned the extra stacked case. The same cannot be said for Korean.

Lee and Ramsey (2000) also point out that *-ka* can even be attached to a non-nominal. This is shown in example (12), where the *-ka* morpheme is attached to the verb *siph* 'want'. It would be hard to accommodate examples such as (11) and (12) if *-ka* and *-lul* can only be used as structural case markers.

(12) *Na-nun ku salam-ul mannako siph-ci-ka anh.a.* [NOM = FOC]
I-NUN that person-ACC meet want-INFL-**NOM** NEG
'I just don't want to see that person.' (Lee&Ramsey 2000, p.146, example (15))

Finally, it appears that *-ka* may be attached to a nominative-case marked nominal, as demonstrated in example (13). In (13), the nominal *sensayngnim* 'teacher' is marked with the honorific nominative case marker and an additional *-ka* marker is stacked upon the honorific marker. Assuming that *-ka* is also a nominative case marker, this double case marking seems redundant and poorly motivated.

(13) *Sensayngnim-tul-kkeyse-man-i kulen il-ul* [NOM = FOC]
teacher-PL-HON-only-**NOM** that.kind work-Acc
ha-si-pnita.
do-HON-Level.Formal.DECL
'Only teachers do such work.' (Sells, 1995, p.294, example (29))

In summary, I have reviewed the evidence that calls for a reanalysis of Korean case stacking as a focus strategy. Cross-linguistically, the Korean case stacking construction differs from its counterpart in other case stacking languages in that only the Korean case-stacked nominal is the head of an independent DP. Language-internally, we have seen differences between the stacked morpheme and the unstacked structural case markers as well as similarities of the former to the Korean focus/topic marker. These all suggest that treating Korean as a case stacking language like Lardil and Warlpiri deserves reconsideration.

3.3 *Case alternation* While Korean should not be considered as a case stacking language, we note that it also should not be attributed to the same category as other non-case-stacking languages either. This is

[5] Unless otherwise cited, all data has been gathered by the author.

because Korean allows case alternation. Compare the Korean example (2) (reproduced here as example (14)) with the German example (15). Both sentences express the same meaning and have similar structures, yet only in Korean can the possessor nominal *Jon* be alternatively marked with the nominative case suffix[6]. I should note that although Wunderlich (2014) reports two alternative translations of the Korean utterance, I am not considering the alternation between the two cases to have structural implications as it would in English.

(14) *Jon-uy/-i son-i cakta.*
John-**GEN/NOM** hands-NOM small
'John's hands are small/John has small hands.' (Wunderlich, 2014, p.347, example (12a))

(15) *John-s hände sind kleine.* [German]
John-GEN hand.PL be small.FEM
'John's hands are small.'

Importantly, we observe the same structural constraint applying to the Korean case-alternating nominal and the case-stacking nominal in Lardil and Warlpiri: the case-alternating nominal has to be embedded in a higher-level DP headed by another nominal. The alternative case in the case alternation construction is also licensed by the structural position of the DP that it is embedded in. This critical observation motivates the case concord analysis that will be discussed in detail in the next section.

4 Case Stacking and Case Alternation: Result of Case Concord

4.1 *Theoretical assumptions* In my analysis, I assume that syntax, morphology and phonology are separate components of the grammar. While abstract case assignment happens within narrow syntax as traditionally conceived, I argue that additional processes that occur in subsequent domains such as morphology and phonology will significantly affect how case is realized in the final output of the grammar.

Divorcing case realization from the narrow syntax might appear to be a bold move, but a similar post-syntactic treatment approach has been adopted by Bobaljik (2006), who argues that agreement between a predicate and its argument is part of morphology rather than (narrow) syntax. In his paper, Bobaljik (2006) points out that morphological case (or m-case) "makes reference to the syntactic structure in their structural description, but they effect no change to the syntactic representation" (p.8). As a result, it is appropriate to place the rules of m-case assignment in the morphological component of the grammar. His position on m-case is essentially what I will adopt here.

4.2 *Case Concord* To reiterate, we have observed that case stacking and alternation constructions are subject to the same structural constraint, i.e. they can only occur on nominals that are embedded in a DP headed by another nominal. This DP in which the case-stacked or case-alternating nominal is embedded essentially determines the additional case in both constructions. Based on this observation, I propose that case stacking and case alternation are both result of case concord, in which the additional case morpheme gets copied onto the case-stacking or case-alternating nominal through a dominance relationship. A similar case concord analysis is found in Nordlinger (1997) and Richards (2013). While these previous analyses focus on using case concord to account for case stacking in Lardil and related Australian languages, I propose

[6] J. Yoon (1996) identifies another form of case alternation in Korean that is closely related to case stacking. According to J. Yoon (1996), case stacking always implies case alternation, as shown in example (1) below.

(1) *Kim sensayngnim$_i$-eykey/eykey-(man)-i/i casin$_i$-uy canglay il-i kekcengsulepta*
Kim teacher-DAT/**DAT-(only)-NOM**/NOM self-GEN future affairs-NOM worry
His$_i$ future worries Prof. Kim. (J. Yoon, 1996, p.112, example (5))

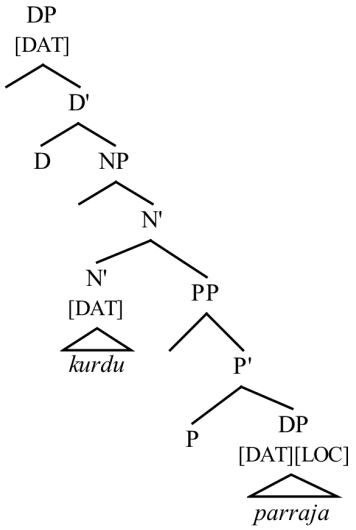
Limited by the scope of this paper, I will not be able to get into the discussion of whether example (1) should be identified as case alternation given my position that case stacking does not exists in Korean. Instead, I will adopt Schütze's treatment of such constructions. As indicated by Schütze (2001), the alternative nominative marking in (1) could in fact be another non-case use of the *-ka* particle in addition to its focus marker role in a case stacking construction. Analogies can be drawn from other double-nominative constructions as well as the ECM constructions in Korean. The reader may refer to Schütze (2001) for the details of this argument.

to extend this mechanism to derive case alternation constructions and thus reach a unified analysis for both case stacking and alternation. This case concord process will happen at the boundary of the syntax and the morphology part of the grammar, as it references the result of case assignment, which is assumed to occur within syntax. Specifically, when case is licensed external to a DP, it is always licensed to the topmost DP projection. Then in morphology, the case percolates downward to all nominals that are dominated by that DP and are eligible for case marking, i.e. nominals occupying the Specifier, Adjunct and Complement positions within that DP[7].

For example, in the case stacking construction (4a) above (partially reproduced in (16)), the indirect object DP consists of the head nominal *kurdu* 'baby' and an adjunct *parraja* 'coolamon', the structure of which is shown in (17). The dative case is licensed external to this DP, presumably by the little *v*, to the highest DP projection. Since the case licensed to the DP is essentially licensed to the head nominal, *kurdu* receives the dative case through case assignment as expected. As the locative adjunct nominal, *parraja* receives a locative case that I assume is assigned by the empty P head in the adjunct PP. In addition, the dative case also percolates from the topmost DP projection to *parraja*, and thereby *parraja* also receives the dative case. The double case morphemes on *parraja* thus have two different sources: the locative case is assigned in syntax while the dative case is copied in morphology.

(16) *kurdu-ku … parraja-rla-ku.*
baby-DAT … coolamon-**LOC**-**DAT**
'baby in the coolamon.'

(17)

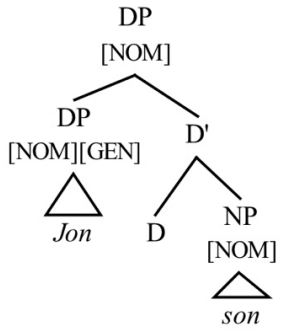


Similarly, for the case alternation construction in example (2) (partially reproduced as example (18)), the subject DP consists of the head nominal *son* 'hand' and the possessive nominal *Jon*, with the structure shown in (19). As nominative case is licensed external to the DP, presumably by Tense, it gets licensed to the topmost DP projection and later percolates to *Jon*. The genitive case, on the other hand, is licensed within the DP to the possessive nominal *Jon.* Again, *Jon* receives the two alternative cases in two different ways: the genitive case through case assignment and the nominative case through case copying.

(18) *Jon-uy/-i son-i*
John-GEN/-NOM hands-NOM
John's hands

[7] The reader might note that one special type of DP is the ones with a relative clause embedded inside. Limited by the scope of this paper, I will not be able to get into the details of how I treat this special construction. I will essentially assume that clausal boundaries block the spreading of case morpheme, but this deserves further research.

(19)



In this case concord analysis, the same process generates the additional case morphemes in both case stacking and case alternation constructions. I make a bolder assumption that this process of case concord happens in all languages, even those that do not show morphological case stacking and/or case alternation such as English. Then the question is, what determines which cases get realized in the surface structure such that we observe case stacking and alternation in some languages but not others? In the next section, I propose a constraint-based model that operates within phonology to control the realization of morphological cases in the output of the grammar.

5 Constraint-based model and language typology

5.1 *Theoretical framework* This constraint-based model of case stacking and alternation adopts the basic theoretical framework of the (classic) Optimality Theory first introduced by Prince and Smolensky (1993). While OT has been used primarily as a framework to study phonology, it has also stimulated important research in syntax, morphology and other areas. For instance, Samek-Lodovici (2001) examines agreement impoverishment under subject inversion in an OT framework, and Woolford (2016) uses a constraint-based model to analyze portmanteau agreement.

I will adopt the three basic components of OT: CON, GEN and EVAL. CON refers to the universal constraint component of the grammar. GEN is the operational component that constructs a set of candidate output forms that deviate from the input. EVAL selects the optimal member of the set of potential outputs to be the actual output (McCarthy, 2007). In the context of case realization, the inputs to GEN are individual constituents with syntactically-assigned and morphologically-copied case features attached to them. These case features are then selectively realized in the eventual outputs.

It should be noted that it is possible to develop an equivalent post-syntactic analysis of case realization without appealing to the OT framework, yet the use of OT can greatly improve the clarity and simplify the analysis. OT is especially powerful when it comes to capturing cross-linguistic variations. As pointed out by Woolford (2007), OT abandons almost all of the common complicated treatment of cross-linguistic differences such as accidental lexical gaps, or language specific rules or principles. Instead, "all cross-linguistic variations in OT must follow from the relative ranking of a set of universal (but violable) constraints" (p.119). In terms of case realization, we observe that languages differ in their treatments of cases that are assigned in syntax and the cases that are copied through case concord in morphology.

5.2 *Constraint-based model and language typology* To account for the different case realization patterns, we first need the faithfulness constraint MAX-CASE that requires the surface form to maximally match the underlying form[8]:

(20) MAX-CASE: All case features should be visible in the output.

[8] MAX-CASE is essentially a stronger form of the MAX(LEXF) constraint proposed by Wunderlich (2008) which requires all lexical case features to be visible in the output.

Meanwhile, languages differ with respect to whether they allow the copied cases to be realized in addition to or as an alternative to the assigned case. As a result, we need two markedness constraints *ASSIGN and *COPY to allow for all of these options:

(21) a. *ASSIGN: Realize all case features except the assigned case.
b. *COPY: Realize all case features except the copied case.

Different relative rankings of these three constraints MAX-CASE, *ASSIGN and *COPY divide languages into at least three typological groups.

Group I is the Lardil type of languages that allow case stacking. In these languages, MAX-CASE should be the highest ranking constraint. Violation of this constraint rules out the candidate outputs that only realize some but not all of the case features that are attached to the constituent in the input. The tableau in (23) shows how case stacking is generated in example (1) (reproduced here as example (22)).

(22) *Ngada latha karnjin-i marun-ngan-ku maarn-ku*
I spear wallaby-ACC boy-**GEN**-**INSTR** spear-INSTR
'I speared the wallaby with the boy's spear.' (Richards, 2013, p.43, example (3))

(23)

marun + [GEN] + [INSTR]	MAX-CASE	*ASSIGN	*COPY
a. ☞ *marun-ngan-ku*		*	*
b. *marun-ngan*	*!	*	
c. *marun-ku*	*!		*
d. *marun*	*!		

Group II is the Korean type of languages with case alternation but not case stacking. To prohibit both the copied and assigned cases from being realized simultaneously, the MAX-CASE constraint should be ranked low in these languages. The *ASSIGN constraint and the *COPY constraint are indistinguishable in their rankings to allow for the alternation. This is demonstrated in the tableau in (25) for example (2) (reproduced here as example (24)).

(24) *Jon-uy/-i son-i cakta.*
John-**GEN/NOM** hands-NOM small
'John's hands are small/John has small hands.' (Wunderlich, 2014, p.347, example (12a))

(25)

Jon+[GEN]+[NOM]	*ASSIGN	*COPY	MAX-CASE
a. ☞ *Jon-uy*	*		*
b. ☞ *Jon-i*		*	*
c. *Jon-i-uy*	*	*!	
d. *Jon*	*	*!	*

Group III is the German type of languages with neither case stacking nor case alternation, but always realize the assigned case when there is an available lexical entry. In these languages, the *COPY constraint strictly outranks both of the other two constraints, as shown in the tableau in (27) for the example sentence in (15) (reproduced here as example (26)).

(26) *John-s hände sind kleine.*
John-GEN hand.NOM.PL be small.FEM
'John's hands are small.'

(27)

John+[GEN]+[NOM]	*COPY	*ASSIGN	Max-case
a. ☞ *John-s*		*	*
b. *John*	*!	*	*

A similar analysis applies to the English example: *I ate the cookie in the jar.* Here, the direct object DP consists of the head nominal *cookie* and the PP adjunct *in the jar.* The accusative case is assigned to the direct object DP and the "locative case" is assigned to the nominal *jar* within syntax. Through case concord, the accusative case gets copied onto the adjunct nominal. Therefore, at this point, the sentence should look like: *I ate* [*the cookie*$_{[ACC]}$ [*in the jar*$_{[LOC][ACC]}$]$_{PP}$]$_{DP}$. Since English belongs to Group III, the *COPY constraint blocks the copied case from being realized on the nominal *jar* and *jar* surfaces with only the "locative case".

While this model is flexible enough to accommodate a fourth or even fifth typological group with a different ranking of the constraints and thus a new case pattern, to the writer's knowledge, there does not seem to exist a language that strictly ranks the *ASSIGN constraint higher than the *COPY constraint (i.e. the opposite of Group III). If such a language were to exist, we would expect a language that looks like example (28) below, where the assigned case is blocked from surfacing.

(28) *Ngada latha karnjin-i marun-(*ngan)-ku maarn-ku*
I spear wallaby-ACC boy-(***GEN**)-**INSTR** spear-INSTR
'I speared the wallaby with the boy's spear.'

Alternatively, a language that does not distinguish between the rankings of these three constraints would allow both case stacking and alternation. Japanese is a possible candidate for such a language (see Nakamura&Fujita 1998, Wunderlich 2014 etc.)[9].

6 Conclusion

In this paper, I reviewed the case stacking and alternation constructions and proposed a unified, post-syntactic approach to account for both phenomena. I argued for a differentiation between case stacking in languages such as Lardil and Warlpiri and pseudo case stacking in Korean. For case stacking and alternation, I proposed that the multiple cases that are attached to a single nominal are either assigned in syntax or copied through case concord in morphology. The result of case concord is then input into the constraint-based model in phonology, in which languages' unique rankings of the three case realization constraints—MAX-CASE, *ASSIGN and *COPY—determine which of the case features are spelled out in the surface structure. This constraint-based model of case realization patterns illuminates a language typology that is both constrained and extendable to accommodate new data.

This research could be extended in multiple directions. First, as previously mentioned in Section 5, I assume that clausal boundaries will block case concord. Yet the exact behavior of nominals embedded within a relative clause or an infinitive clause that modifies a head nominal deserves further investigations. How such nominals are treated in the case concord analysis should be specified. It has also been pointed out to me that the *ASSIGN constraint proposed in the case realization model can be better grounded. Future research will involve collecting additional, independent evidence to support the proposal of the *ASSIGN constraint. Finally, research effort can also be devoted into completing the language typology and analyzing data of languages that belong to the other typological groups that are not discussed in detail in this paper.

References

Bobaljik, Jonathan David. (2006). Where's Phi? Agreement as a post-syntactic operation. In M. van Koppen (Ed.), *Leiden papers in linguistics* (pp.1–23). Leiden: Universiteit Leiden.

Chomsky, Noam. (2000). Minimalist inquiries: The framework. In R. Martin et al. (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (pp.89–155). Cambridge, MA: MIT Press.

Chomsky, Noam. (2001). Derivation by phase. In M. J. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp.1–52). Cambridge, MA: MIT Press.

[9] It has been brought to my attention that as the model currently stands, cases that are assigned in syntax and cases that are copied in morphology are not distinguished in the input to the constraint-based model. I propose it is possible that copied cases are marked with a COPY subscript as they get attached to the target nominals in case concord. The idea of attaching a marker or diacritic to certain morphemes to differentiate them from some otherwise indistinguishable morphemes has been used to account for Opacity in OT Phonology (see Green (2004), Sanders (2003) etc.). I leave the details of this proposal for future research.

Gerdts, Donna. B., & Youn, Cheong. (1988). Korean psych constructions: Advancement or retreat? In L. MacLeod, G. Larson and D. Brentari (Edls.), *Papers from* the *24th Annual Regional Meeting of the Chicago Linguistic Society, Part One: The General Session* (pp.155–175).

Green, Anthony D. (2004). Opacity in Tiberian Hebrew: Morphology, not phonology. In *ZAS Papers in Linguistics, 37*, 37-70.

Lee, Iksop, & Ramsey, S. Robert. (2000). *The Korean Language.* New York, NY: State University of New York Press.

Lees, Aet. (2015). *Case Alternations in Five Finnic Languages: Estonian, Finnish, Karelian, Livonian and Veps.* Leiden: BRILL.

Levin, Theodore. (2014). Successive-cyclic case assignment: Korean nominative-nominative case-stacking. Abstract for 9th Workshop on Altaic Formal Linguistics.

Martin, Samuel Elmo. (1992). *Reference Grammar of Korean: A Complete Guide to the Grammar and History of Korean Language.* Rutland, VT: Charles E. Tuttle.

McCarthy, John J. (2007). What is Optimality Theory? in *Language and Linguistics Compass.* 93.

Nakamura, Wataru. (1997). *A Constraint-Based Typology of Case Systems* (Doctoral dissertation). SUNY, Buffalo, NY.

Nakamura, Wataru, & Fujita, Takeshi. (1998). Case Alternations in Potential Constructions in Japanese and their Semantic Implications. *Language, Information and Computation (PACLIC 12),* 172-183.

Nordlinger, Rachel. (1997). *Constructive Case: Dependent-marking nonconfigurationality in Australia* (Doctoral dissertation). Stanford: Stanford University.

Prince, Alan, & Smolensky, Paul. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar.* In Rutgers University Center for Cognitive Science Technical Report 2. Piscataway: Rutgers University Center for Cognitive Science.

Richards, Norvin. (2007). Lardil 'Case Stacking' and the Structural/Inherent Case Distinction. *Syntax, 16,* 42-76.

Samek-Lodovici, Vieri. (2002). Agreement impoverishment under subject inversion – A cross-linguistic analysis. In G. Fanselow and C. Féry (Eds.), *Resolving conflicts in grammar: Optimality Theory in Syntax, Morphology, and Phonology. Linguistiche Berichte Sonderheft*, *11,* 49-82.

Sanders, Robert Nathaniel. (2003). *Opacity and Sound Change in the Polish Lexicon* (Doctoral dissertation). University of California, Santa Cruz.

Schütze, Carson T. (2001). On Korean 'Case Stacking': The varied functions of the participles *ka* and *lul*. *The Linguistics Review, 18,* 193-232.

Sells, Peter. (1995). Korean and Japanese morphology from a lexical perspective. *Linguistic Inquiry, 26,* 277-325.

Simpson, Jane. (1991). *Warlpiri morphosyntax: A lexicalist approach*. Dordrecht, The Netherlands: Kluwer.

Vermeulen, Reiko. (2012). Word order variation and information structure in Japanese and Korean. In J. Koster et al. (Eds.), *The Syntax of Topic, Focus and Contrast: an Interface-based Approach. Studies in Generative Grammar, 133*, 77-118.

Woolford, Ellen. (2007). Introduction to OT Syntax. *Phonological Studies, 10,* 119-134.

Woolford, Ellen. (2016). Two Types of Portmanteau Agreement. In G. Legendre, M. Putnam, and E. Zaroukian (Eds.), *Advances in OT Syntax and Semantics* (pp.111-135). Oxford: Oxford University Press.

Wunderlich, Dieter. (2008). The force of lexical case: German and Icelandic compared. In K. Hanson and S. Inkelas (Eds.), *The nature of the word: Essays in honor of Paul Kiparsky* (pp.587-619). Cambridge, MA: MIT Press.

Wunderlich, Dieter. (2014). Variations of double nominative in Korean and Japanese. *Studies in Language and Cognition*, Düsseldorf, Germany: Düsseldorf University Press.

Yoon, James H. (1996). Ambiguity of government and the Chain Condition. *Natural Language and Linguistic Theory, 14,* 105–162.

Yoon, James H. (2004). Non-nominative (major) subjects and case stacking in Korean. In P. Bhaskararao and K. Venkata Subbarao (Eds.), *Non-nominative subjects*, vol. 2, pp.265–314. Amsterdam: John Benjamins.

"It's More Complicated than That": Women's Constructions of Non-Normative Sexual Identity

Christopher Dale
Boise State University

1 Introduction

When it comes to sex and sexual identity, all cultures hold specific ideologies, in particular when it comes to women. In contemporary U.S. culture, many have become familiarized with the concept that sexual identity, at its core, exists on a fluid spectrum wherein a person can identify in one particular way for a set period of time before moving on to utilize different identity labels. Most commonly, we see this unidirectionally – for example, a man identifying as straight before "coming out of the closet" and identifying with more non-normative labels, such as gay or bisexual, only rarely returning to identify with heterosexuality (this has been discussed ad nauseum in the literature; see Downs 2012 for one example). However, the phenomenon of sexual fluidity has been broadly conceptualized culturally as applying mostly to women. That is to say, women are more likely to be portrayed or allowed by the culture to explore the concept of sexual fluidity than their male counterparts. From news sites like The Guardian (Theobald 2013), Huffington Post (Chapin 2014), and CNN (Kerner 2012), to television shows like Netflix's internationally acclaimed *Orange is the New Black* (Walsh 2013), it would almost seem like there is an expectation for women to have a sexually fluid identity in today's world.

This notion of a sexually fluid woman is pushed further by other ideologies that exist in the culture which claim, for example, that women are more likely to have had a partner of the same-sex and that women are more likely to have engaged in sexual experimentation with a member of the same-sex. It would appear, however, that there is some truth to these ideologies. A recent report published by the CDC that documented a survey they conducted between 2011 and 2013 suggests a statistically significant variation between men and women in terms of sexual practices. In this report, 17.4% of the female responders confirmed having had same-sex sexual contact at some point in their life, in comparison to only 6.2% of the males (Copen 2016). Significantly, these numbers do not directly correlate to women (or men) who identify as something other than heterosexual. In fact, the majority of respondents overwhelmingly identified as either heterosexual or straight, with 92.3% of women and 95.1% of men. If one were to interpret why this gap between experience and identity exists, one could point to the prevalence, or perhaps more accurately, the dominance of heterosexuality within the culture.

Inspired by recent work in the field on men's non-normative sexual experiences and the ways in which they coincide with the idea of hegemonic heterosexuality (VanderStouwe 2016), this paper attempts to address a lack of sociolinguistic research on the topic of women's negotiation of sexual identities and the ways in which they are discursively constructed when women have had non-normative sexual experiences. Through the use of social problem oriented conversation analysis (Kasper and Wagner 2014), this paper analyzes women's discourse regarding their sexual identities in an attempt to discover how women apply identity labels to fluid sexuality.

In beginning the research that lead to this paper, I first suggested that women would choose sexual identity labels that allowed space for alternative sexual experiences, such as sexually fluid, "no label", or the increasingly popular and equally contentious "queer" (Cho 2013). However, knowing that society expects women to return to heteronormativity after a brief foray into atypical, non-normative sexual experience, I hypothesized that women who have had said experiences, when asked, would identify in much more concrete and typical ways than what the ideologies about their sexual fluidity would suggest. I hypothesized that these women would have an intimate understanding of their own sexual identity and would have ways in which they could express it. It is from this hypothesis that I conducted and approached my research.

Proceedings of *WECOL 2017*

2 Background

A heterosexual orientation is expected for both women and men – that is the predominate sexual identity throughout much of Western culture (Kitzinger 2005; Land and Kitzinger 2005). While men may struggle to be accepted for deviations to this paradigm and go through efforts to avoid the public eye if exploring non-heterosexual encounters (VanderStouwe 2016), for women, this phenomenon is discursively constructed to provide acceptability for a set time period of sexual experimentation. Popular concepts such as the "L.U.G." (Lesbian Until Graduation) demonstrate how terms of temporality used to classify and limit the sexual expression of women are generated by the culture (Lewin 2011). It's acceptable, even preferable, for a woman to exhibit and identify as anything other than heterosexual, as long as it's "just until graduation".

In this way, because there is such a different and nuanced cultural expectation for the sexuality of men and women, gender and sexuality are inextricably linked (Valentine 2004). In other words, it is difficult to interpret the discursive interactions of men and women as belonging to one large, group (i.e. "male discourse" or "female discourse") without considering other facets of the speaker's identity, such as their sexuality. To provide an example, both gender and sexual identity have played a significant role in linguistic research regarding the discursive tools female, heterosexual speakers are able to utilize in conversation regarding their sexual practices (Jackson and Cram 2003); without an analysis of both gender and sexuality, the observations made by Jackson and Cram would be difficult to interpret in a broader cultural context.

To further the idea of the inseparability of different facets of an individual's identity, Levon (2015) brings the theory of intersectionality into the field of language, gender, and sexuality research. Like Crenshaw (1989), Levon argues that the various aspects of one's identity cannot be separated out into simple categories such as "male language" or "gay language", because there are other factors that influence one's linguistic background, like race, class, or ability. In putting forth this argument, Levon hopes that researchers will take a speaker's full identity into account before drawing generalizations that may not apply universally to all potential speakers within a given identity category. With this in mind, it becomes increasingly important to not assume that women of differing backgrounds, despite having had similar experiences, will have analogous methods of dealing with discursively constructing their own sexual identities to those of men. Instead, research must be conducted to discover the linguistic tools that the women employ.

Taking the theory of intersectionality one step further, it becomes evident that language use is not only influenced by one's own, claimed identity, but by the ideologies held and the actions practiced by the speaker (Bucholtz and Hall 2004). Applying this understanding to sexual identity, and specifically, women's sexual identity, avenues for analysis open up that include not only the language a speaker employs to talk pointedly about their identity, but the ways in which those employments are complicated and given significance by prevalent ideologies and specific sexual practices that the speaker has experienced. For instance, speakers who have had non-normative sexual experiences, but who do not claim an identity that is in alignment with said experiences, will logically have to employ some sort of discursive tool in order to have their identity be understood by interlocutors. This theory is crucial in interpreting the data collected for this paper, as it is directly reflected in the research.

3 Methodology

Data for this paper comes from recordings conducted at a large public university in the Pacific Northwest. In order to participate in the study, each participant had to meet a set of five specific criteria. Two of these were most important in terms of my research: 1) participants had to currently identify as being a woman; and 2) participants had to have had at least one non-normative sexual experience. Non-normative sexual experiences were further defined as: 1) the participant had been physically intimate (kissing, sexual intercourse, etc.) with a member of the same-sex; and 2) the subject continues to experience desire to engage in these activities. The other three criteria were used to limit the pool of potential participants, and dealt with geographic location, age, and enrollment in university. These criteria were used to limit participants to those

born or raised primarily in the Northwestern United States, those who were between the ages of 20 and 25, and those who were currently enrolled at a university.

At the time of data collection, each participant attended the same university; was between the ages of 20 and 23; was a woman who had had non-normative sexual experiences; and for whom all, except one, were born and raised in the Pacific Northwest. The one participant who was not raised in the Pacific Northwest had however spent most of her life living in the area. Additionally, all of the participants knew each other before the recordings took place. All participants were of similar socioeconomic background and were white. The participants who appear in the data that follows include: Sara, a 20-year-old Social Science major; Ana, a 21-year-old Foreign Language major; Zenya, a 21-year-old Fine Arts major; and Liz, a 23-year-old Fine Arts major. Due to the size of the university which they attend and the sensitive nature of the research topic, their names have been anonymized and their fields of study generalized.

Research was conducted via two in-person interviews, lasting around an hour each. Each of the interviews featured two of the participants – Sara and Ana in the first, and Zenya and Liz in the second – in conversation generated from questions that I posed. In order for the data to be representative of the women's actual thoughts and feelings, I tried to stay somewhat distant from the conversation so that responses reflected more of the experiences of the women, and so they weren't influenced by my input. Furthermore, each of the participants were familiar with each other and myself prior to entering the research setting.

Of the nearly two hours of data I collected, the portion of the interview presented in this paper focuses on a discussion prompted by the questions, "How do you identify?" and "Why do you identify that way?". This segment of the interview was chosen primarily because while most of the data collected would be relevant in terms of subject matter, it is this conversation in particular that most directly addresses my hypothesis. It is my hope that this discussion could serve as a guide for both myself as well as other researchers looking to conduct further research regarding this topic. The data was analyzed qualitatively, using social problem oriented conversation analysis (Kasper and Wagner 2014), and the transcription provided is based off of the Discourse Transcription system (Du Bois et al. 1992).

4 Data Analysis

In order to explore these women's identity constructions as presented in the interview data collected, three separate qualities of speech that appeared most frequent within my data will be discussed: (1) the use of discourse markers indicating uncertainty in discussing their identity labels; (2) pragmatic arguments for or against the use of identity labels; and (3) phonological markers, such as extended prosody, pauses, and repetitions.

Perhaps the most obvious of these qualities is the reluctance of the women to choose a label when it came to their sexual identity. This occurred most frequently in conversational turns where the participants argued equally against the use of the normative label *straight* and more non-normative labels, like *bisexual* or *pansexual*. This is illustrated in Example 1, featuring Ana's response to the question, "How do you identify?":

(1) "I like don't even know"

22 ANA; I like,
23 don't even know,
24 I feel like every time I think about it,
25 I'm like in my head,
26 like <FALSETTO> oh I'm straight </FALSETTO>,
27 but then,
28 I'm just not @@,
29 I don't know,
30 but I guess I wouldn't I identify like as ^bisexual,
31 but not like ^pansexual either,
32 so I guess it's just like whoever I find,
33 that I like,

34 [no matter] which gender,

As seen in this example, on line 26 Ana utilizes a change in voice quality, from modal to falsetto, in order to mimic the voice in her head while she plays with the idea that she might be straight. This change in voice quality gives the implication of an almost comical effect, as if it's not actually herself telling her that she's straight or as if she doesn't fully believe this to be true. This is further supported by her claim on line 28, where she says, "I'm just not" in reference to her identifying using the label "straight". While this distinction could be seen as a way to definitively align her sexual identity with one of the non-normative sexual identity labels, in lines 30 and 31, she similarly rejects the label "bisexual" and "pansexual" without providing any sort of reasoning as to why these labels do not fit her experiences. Ana does not name a concrete label with which she identities her sexuality. Instead, she discursively constructs her identity by arguing against labels she does not claim for herself. Then, in her own terms, she describes in detail how she identifies. From lines 32 to 35, she describes her sexual identity as "whoever [she] finds that [she] likes," no matter what their gender might be.

Similarly, Sara was also hesitant to use label "bisexual", but she did not mention her relationship to the label "straight". After discussing some of the preconceptions and ideologies those close to her hold towards bisexuality, she described her identity in example 2, as follows:

(2) "I don't like, want to use that word"

6 SARA; I guess I don't really like,
7 ever say what I identify as,
8 like,
9 like I don't want to say bisexual,
10 cause that just like,
11 like my parents were always like,
12 bisexuals are [slutty],
(3 lines omitted)
16 SARA; you know,
17 like I've never thought that,
18 but I don't like,
19 want to use that word,
20 but I ^guess that is what it is @@.

Sara demonstrates the same sort of resistance to preexisting labels as Ana did in the first example. In line 19, Sara refers to the label "bisexual", and like Ana, she says that she doesn't "want to use that word" to identify herself. Unlike Ana, however, Sara never makes an argument against using the label "straight" to refer to herself. Instead of creating an identity by arguing against what she is not, Sara finally concedes to the label of "bisexual" and says, "I guess that is what it is", in line 20.

With this example, a second quality of speech emerged that I was able to trace across each of the participants' responses: whenever identity was being discussed, specifically when it was about the personal identities of the women involved in the study, discourse markers indicating uncertainty almost always punctuated conversational turns. In example 2, Sara eventually concedes and says that the word "bisexual" would be a term that could be used to describe her own sexual identity (line 9), but it is prefaced with the words "I guess" (line 6), further underlining the uncertainty that she has in regard to this identity label. Turning back to the first example, Ana also marks aspects of her conversational turn with discourse markers indicting uncertainty. However, unlike Sara, she begins her turn with one of these discourse markers. In line 23, her first reaction to the question is to say, "I don't even know." She then goes on to mark both her rejection of preexisting labels (line 30, "but I guess I wouldn't identify as like ^bisexual") and her description of her own identity (line 32, "so I guess it's just like whoever I find") using similar discourse markers.

These discourse markers indicating uncertainty even show up in the speech of the only participant to concretely define her sexuality using established labels, as seen in Zenya's response in Example 3.

(3) "Bisexual, possibly pansexual"

89 ZENYA; Um::,
90 yeah,
91 bisexual,
92 possibly pansexual,
93 I guess they,
94 they're kind of interchangeable,
95 I feel.

Despite the initial hesitation (lines 89-90), which could be explained as a natural pause in the conversation, Zenya is quick to identify herself as bisexual, or "possibly pansexual" (lines 91-92), as she finds both terms interchangeable. This denotes that both terms are likely used to describe her sexual identity. Unlike the other participants, she has a clearly marked identity that she describes using terms that are readily accessible and understood by most interlocutors. However, when the women were asked why they identified in the ways they discussed, Zenya had a different response, shown below in example 4.

(4) "That's a really hard question"

121 ZENYA; I think,
122 I identif::y,
123 you ask me why I like ^girls @@?,
124 I::,
125 that's a really hard question.
(15 lines omitted)
140 ZENYA; Cause I'm sexually attracted to both sexes,
141 well,
142 I guess all sexes,
143 all genders,
144 bu::t .

Zenya's responses in the two separate conversational turns in example 4 demonstrate a remarkably different speech style than seen in example 3. Here, she expresses a level of uncertainty and hesitation that was not present in her first response. In line 142, like Ana and Sara in the examples 1 and 2, Zenya marks the description of her identity with a discourse marker indicating uncertainty, "I guess". While here this statement seems to be serving more of a corrective function than the previous examples, another similar discourse marker, "I think", appears on line 121. Rather than serving as a statement of personal opinion, her use of "I think" can instead be seen as a sign of uncertainty. This understanding of the phrase, "I think" is bolstered by the elongated prosodies found in the final syllable of the word "identify" in line 122 and again on the personal pronoun "I" in line 124. In addition, line 125 reveals the lack of a specific response to the question until her next conversational turn (lines 140-144), after Liz had described her reasoning for choosing her own sexual identity label.

With example 4 in particular, there is strong evidence for the third and final quality of speech I noticed in my data. That is, while there may not be as clear and evident of a use of discourse markers indicating uncertainty in example 4 as there were in examples 1 and 2, Zenya utilizes phonological markers that seem to serve a similar discursive function. This is shown through the use of extended prosody as exemplified by the words "identify" on line 122, "I" on line 124, and "but" on line 144, and additionally through the pause at the end of her second conversational turn on line 144.

Similarly to Ana and Sara's usage of the term "I guess" (Example 1, Lines 30, 32; Example 2, Lines 6, 20), Zenya uses extended prosody and pauses to show that while she is currently certain of the labels she's chosen to identify herself, she is uncertain as to why she sees those labels as adequately encompassing her non-normative sexual experiences. Zenya is not the only participant to use these phonological markers for this particular function, however. Liz also uses extended prosody and pauses to denote uncertainty towards

the topic of her own sexual identity. In fact, out of all the participants, these phonological markers appear most evidently in Liz's speech, particularly in her response to the question, "How do you identify sexually?" in example 5.

(5) "I have no idea"

96 LIZ; I have no idea,
97 for the,
98 for,
99 for the most part,
100 I,
101 I::,
102 I identify as heterosexual,
103 but there's parts of me that are like,
104 well maybe not,
105 (..)
106 so maybe like bi-curious.

In this example, the most evident phonological feature is repetition of certain words. For example, in lines 97-99, the word "for" is repeated three times in the restarting of her intonation unit. In lines 100-102, the word "I" is similarly repeated three times before a full intonation unit is completed. Additionally, there is extended prosody on the second "I" in line 101. Applying what was evidenced in the other examples, particularly in Example 4 where Zenya repeatedly employs extended prosody in an attempt to explain why she has chosen bisexual/pansexual as her sexual identity label, we can infer some particularly interesting things about Liz's usage here – namely, that her employment of extended prosody and repetition further exemplifies the uncertainty she feels in the telling of her identity label.

To analyze this example further, similarly to Ana's use of falsetto on lines 25-28 in example 1, Liz states in example 5 that she identifies as heterosexual in line 102 before immediately qualifying it in lines 103 and 104 with "but there's parts of me that are like, well maybe not". This signifies the recalling of an internal dialogue where Liz wrestles with what to properly call herself in order to encapsulate her non-normative sexual experiences within her sexual identity label. As seen with Zenya, this is further demonstrated through Liz's usage of extended prosody and repetition, which discursively functions here like the use of discourse markers indicating uncertainty found in previous examples. Furthermore, on line 106, Liz uses a similar discourse marker to preface what she would currently call herself within the context of the study: "so maybe like bi-curious". Throughout this example, it is evident that Liz's usages of extended prosody and repetition mirror the usages of these markers by other women in the study, as there are similar implications which can be found across each of their speech styles.

In addition, even the label "bi-curious" itself denotes a level of uncertainty or hesitation to claim the label as one's own. It signifies someone who is curious about identifying as bisexual – someone who wants to try the label out but doesn't yet want to fully claim it for themselves. Moreover, Liz's claim to the label of bisexual, as evidenced in this example, is much more similar to Sara's claim to it than it is to Zenya's. Unlike Zenya, Liz does not immediately claim the label for herself but begins by first explaining how she identifies, then attempting to place a label onto it. She ends her conversational turn by saying, "maybe bi-curious", in a way that mirrors Sara's "but I guess that's what it is" found on line 20 of the example 2, in that it is not necessarily by her own choice that she chooses to use this label, but because it is a label that can be easily understood by interlocutors. In other words, the label doesn't have to completely embody Liz's sexual identity, as long as it conforms to some sort of general understanding of her sexual experiences.

5 Discussion

The hypothesis that women who have had non-normative sexual experiences would identify in concrete ways, such as through using established sexual identity labels, was disproven. The research suggested that, generally, these women were hesitant to label themselves, or at the very least, uncertain as to what label to

use. Rather, the women in the study used pragmatic arguments against these established identity labels to form identities outside of the bounds of conventional sexuality labels. They aren't straight, but they aren't bisexual either. Instead, the majority of these women used an "unlabel" – that is, the women generally didn't choose specific labels to identify themselves, but instead chose to describe their experiences and their feelings towards their sexual identity, thus forming a more fluid sexual identity through their discourse.

In the one example where a definitive sexual identity label was named as representative of the participant's sexuality (Example 3, Lines 91-92), further questioning on the selection of that label proved to be equally contentious. The reasoning behind the selection of that particular label aligned with the reasoning why the other women in the study chose not to select a label. In addition, the conversational turns describing the reasoning were similarly punctuated by discourse markers indicating uncertainty, such as "I guess" (Example 4, Line 142), and the same phonological markers, like extended prosody (Example 4, Lines 122, 124, 144). This correlation points to a universality in the understanding of non-normative sexual experiences by the women who have them. Even though it is difficult to map on significant communicative interpretations of these experiences, the women all struggle in similar discursive ways.

Similarly to what popular ideologies within the culture suggest, this research illustrates that the women who have had non-normative sexual experiences were more likely to have an abstract or fluid sense of their sexual identity. While this discovery could confirm the ideology behind concepts such as the "L.U.G.", the data examples above reveal this unlikely to be the case, particularly because of the way in which each participant described their sexual identity. Each of the women described their identities as being to whoever they are attracted to, regardless of gender. With ideologies like the "L.U.G.", women are expected to identify purely within the confines of one particular sexual orientation. In the case of the "L.U.G.", the women would identify as lesbians and would not hesitate to call themselves as such. If the women were simply experimenting, on the other hand, they might choose identity labels that express their desire for experimentation to interlocutors, like questioning. That the women chose neither of these options necessitates an analysis that goes deeper than comparing the women's discourse to preeminent ideologies. An analysis that points to a sense of shame felt by the women in identifying as something other than heterosexual can equally be ruled out, because while the geographical area that this study was conducted in is fairly conservative and homophobia is still rampant in the broader U.S. culture, the responses seemed to be prompted much more by a lack of appropriate terminology than from a lack of comfortability with the women's lived experiences.

As in VanderStouwe (2016), the discourse markers indicating uncertainty and the extended prosodies, repetitions, and pauses expressed by participants could be due in large part to the participant's navigation of a complex interaction between ideology, practice, and identity. In this way, many of the terms that exist in the English language do not adequately account for the participant's lived experiences. And while there are certain identity labels that could be used, such as bisexual or pansexual, the indexicality of these labels does not align with how the women see themselves in terms of sexual practice. This can be further interpreted as being rooted in ideologies surrounding established non-normative sexual identity labels that would suggest something about the women's sexual practice that does not coincide with their lived reality – that they are sexually promiscuous, for example (Example 2, Lines 11-12).

Unlike in VanderStouwe (2016), however, there is little evidence in this study to suggest that the women avoid selecting identity labels for their sexualities on the basis of desire alone. The women did not construct a sexually fluid identity in order to appear more available to potential sexual partners – quite the opposite occurred. In conversations I had with the women outside of the research context, each of them expressed an inadequacy they felt in approaching potential female partners because of the ways in which they discursively identified their sexuality. Some of the women said that they felt nervous approaching other women because they wanted the woman they were interested in to know that they were serious about their feelings, and that this wasn't going to just be an experiment to "test things out". Furthermore, all of the women in the study said that they did not find it at all desirable to be wanted sexually by men because they had had sexual experiences with other women. In fact, all of the women said that they would actively avoid having sexual encounters with men who had expressed interest in the fact that they had had same-sex sexual experiences.

Ruling out increased sexual desirability as a possible indicator for the appearance of the linguistic hesitations and uncertainties exemplified through the women's discourse, the research points to another source: ideological conflicts. Based on the current data, it can be concluded that there is an ideological

conflict that exists between established identity labels that should account for the speaker's sexual practices on a purely semantic basis, but due to extenuating ideologies surrounding these labels, cannot be used to adequately describe the speaker's sexual identity. This is typified in the research via the women's negotiations of established identities and their struggles to claim those labels as their own (e.g. Example 1, Lines 30-31; Example 2, Lines 17-20). Beyond the context of this study, more research would need to be done to see if this is generally true about women who have had non-normative sexual experiences, and to provide some information as to if and why this conflict exists solely for women.

6 Conclusion

Though there has been recent inquiry into how male speakers navigate sexual identity to incorporate non-normative sexual experiences (VanderStouwe 2016), there has been little sociolinguistic research done to investigate the ways in which female speakers discursively construct their identities to incorporate their own non-normative sexual practices. This paper attempts to fill this gap by using social problem oriented conversational analysis (Kasper and Wagner 2014) and intersectionality theory (Levon 2015) to analyze the discourse of four female participants. While it was initially posited that the participants would have concrete understandings of their sexualities, which would be demonstrated evidently in the linguistic space (i.e. through the use of established labels), this study showed that the women had more fluid conceptions of their identities.

Unlike in previous studies (VanderStouwe 2016), the participants did not utilize fluid constructions of their sexual identities in order to increase their desirability to potential sexual partners. Rather, the women discursively constructed their identities in an attempt to maneuver around complex ideologies surrounding established labels. These ideologies presented a conflict wherein the participants might otherwise use an established label (i.e. bisexual), as they encapsulate the speaker's sexual practices on a basic semantic level, but because these ideologies are so prevalent within the culture, the established labels suggest aspects of the participants' sexual practices that do not match with their lived reality. Thus, the established labels cannot be used to adequately capture the women's sexual identities.

Moving forward, further research in this area would be helpful to draw more concrete and general conclusions about what has been elucidated here. Future research should include a widened sample size so that the discourse of a broader variety of women with non-normative sexual experiences can be analyzed. In this way, the results would be more reflective of a larger speech community of women. Although I do not believe this was a problem in my own research, additional research could also be conducted by a woman with similar experiences so that the participants can feel more connected to the researcher and thus more comfortable with sharing intimate details about their sexual history and practices. Finally, if a similar study were to be conducted featuring men as the participants, we would find elements of different discursive styles that, in tandem with what we are learning about women and their constructions of identity, would broaden our understandings of how sexuality and sexual identity labels function in the 21st century.

7 References

Bucholtz, Mary and Hall, Kira. (2004, Sep.). Theorizing Identity in Language and Sexuality Research. *Language in Society, 33(4)*, 469-515.

Chapin, Angelina. (2014, March 8). *Why Men Are Less Sexually Fluid Than Women.* Retrieved from: http://huffingtonpost.ca/angelina-chapin/sexual-fluidity_b_4538607.html

Cho, Nadia. (2013, June 29). *Being Queer Means...* Retrieved from: http://huffingtonpost.com/nadia-cho/being-queer-means_b_3510828.html

Copen, Casey E., Chandra, Anjani, & Febo-Vasquez, Isaedmarie. (2016, Jan. 7). Sexual Behavior, Sexual Attraction, and Sexual Orientation Among Adults Aged 18–44 in the United States: Data from the 2011–2013 National Survey of Family Growth. *National Health Statistics Reports, 88.* Retrieved from: https://cdc.gov/nchs/data/nhsr/nhsr088.pdf

Crenshaw, Kimberlé Williams. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *The University of Chicago Legal Forum, 140,* 139-167.

Downs, Alan. (2012). *The Velvet Rage: Overcoming the Pain of Growing Up Gay In a Straight Man's World.* Boston, MA: Da Capo Press.

Du Bois, John W., et al. (1992). Discourse Transcription. *Santa Barbara Papers in Linguistics, 4.*

Jackson, Sue and Cram, Fiona. (2003, Mar.). Disrupting the sexual double standard: Young women's talk about heterosexuality. *British Journal of Social Psychology, 42,* 113-127.

Kasper, Gabriele and Wagner, Johannes. (2014). Conversation Analysis in Applied Linguistics. *Annual Review of Applied Linguistics, 34,* 1-42.

Kerner, Ian. (2012, Feb. 9). *Understanding Females' Sexual Fluidity*. Retrieved from: http://thechart.blogs.cnn.com/2012/02/09/understanding-females-sexual-fluidity/

Kitzinger, Celia. (2005). Heteronormativity in Action: Reproducing the Heterosexual Nuclear Family in After-hours Medical Calls. *Social Problems, 52(4),* 477-498.

Land, Victoria and Kitzinger, Celia. (2005). Speaking as a Lesbian: Correcting the Heterosexist Presumption. *Research on Language and Social Interaction, 38(4),* 371-416.

Levon, Erez. (2015). Integrating Intersectionality in Language, Gender, and Sexuality Research. *Language and Linguistics Compass, 9(7),* 295–308.

Lewin, Tamar. (2011, March 17). *Study Undercuts View of College as a Place of Same-Sex Experimentation.* Retrieved from: http://nytimes.com/2011/03/18/education/18sex.html

Theobald, Stephanie. (2013, Nov. 26). *Lesbianism: Sexual Fluidity is a Fact of Life for Women*. Retrieved from: https://theguardian.com/lifeandstyle/2013/nov/26/lesbianism-women-sexual-fluidity-same-sex-experiences

Valentine, David. (2004). The Categories Themselves. *GLQ: A Journal of Lesbian and Gay Studies, 10(2),* 215-220.

VanderStouwe, Chris. (2016, December). *"Straight-ish": Constrained Agency and the Linguistic Constructions of Sexual Identities, Desires, and Practices among Men Seeking Men*. Retrieved from: https://www.academia.edu/30313830/_Straight-ish_Constrained_Agency_and_the_Linguistic_Constructions_of_Sexual_Identities_Desires_and_Practices_among_Men_Seeking_Men

Walsh, Candace. (2013, Oct. 5). *Sexual Fluidity and Orange Is the New Black*. Retrieved from: http:// huffingtonpost.com/candace-walsh/sexual-fluidity-and-orang_b_3701933.html

Uto-Aztecan Lexicostatistics 2.0

Jason D. Haugen[1], Michael Everdell[2], and Benjamin A. Kuperman
[1]*Oberlin College* and [2]*The University of Texas at Austin*

1 Introduction

Uto-Aztecan (hereafter UA) is one of the largest language families of the Americas in terms of geographic spread, number of languages, and number of speakers. The family's overall unity has been known for some time, being definitively established when Sapir (1913, 1915) united the Aztecan and Shoshonean (Numic) languages. However, there is still significant controversy over the internal classification of the UA subgroups. J. Hill (2011) reviews the evidence for sub-groups based on shared phonological innovations but the jury may still be out for a few more groups based on possible shared morphological innovations (Haugen *forthcoming*). In addition to a century of work using the classical comparative method, the UA family has also had a long history of quantificational phylogenetic studies. Lexicostatistics, where subgrouping is based on quantitative analysis of shared cognates, was used in Miller (1984) (who sub-grouped using a metric based on "cognate density"), Cortina-Borja & Valiñas (1989), and Haugen, Everdell & Kuperman (2015, 2016) (who update Miller's study by using "relative cognate density" as a metric). Miller et al. (1971) use glottochronology to propose time distances for diversification of languages within the Numic sub-group. More recently, the Automated Similarity Judgment Program (Holman et al. 2008) and Wheeler & Whiteley (2014) approach the UA subgrouping problem using phonostatistics, which relates languages by phonological similarity measurements. Finally, Cortina-Borja et al. (2002) combine lexico- and phonostatistics to assess the UA subgrouping.

Our goal here is to present the initial results from some new quantitative approaches towards UA subgrouping, by comparing the results of several lexicostatistical clustering methods and wordlists. In the following section we discuss the background of the UA family and major questions in UA family-internal subgrouping. Then, in §3 we discuss the new data that we are using to further the discussion of UA phylogeny. In §4 and §5 we present our results of analyzing UA phylogeny through distance-based and Bayesian methods, respectively. Finally, §6 wraps up with a discussion of our broad-scale findings.

2 Background

Figure 1 shows the geographic spread of the UA language family, and Figure 2 shows the languages we have included in this study, organized into the generally accepted subgroups. The major open questions in UA sub-classification are the existence of the Northern or Southern Uto-Aztecan nodes—these geographical designations have also been claimed to be genetic. For example, Heath (1977, 1978) and Manaster Ramer (1992) find evidence for shared phonological and morphological innovations that suggest a Northern Uto-Aztecan (NUA) node. Southern Uto-Aztecan (SUA) has largely been defined by lexicostatistical evidence, though Merrill (2013) argues for shared phonological innovations. Within NUA, J. Hill (2011) argues against the traditional Takic node and posits instead a Californian subgroup joining the Takic languages (split into Serran and Gab-Cupan) to Tübatulabal (traditionally regarded as a UA singleton). Within SUA, Miller (1984) posits a Sonoran subgroup including the all of the southern subgroups except Aztecan. The place of Tubar remains a critical open question, as that language has been shown to have affinities with multiple SUA groups including Taracahitan, itself called into question by J. Hill (2001), and Tepiman (Stubbs 2000).

Proceedings of *FUAC 2017*

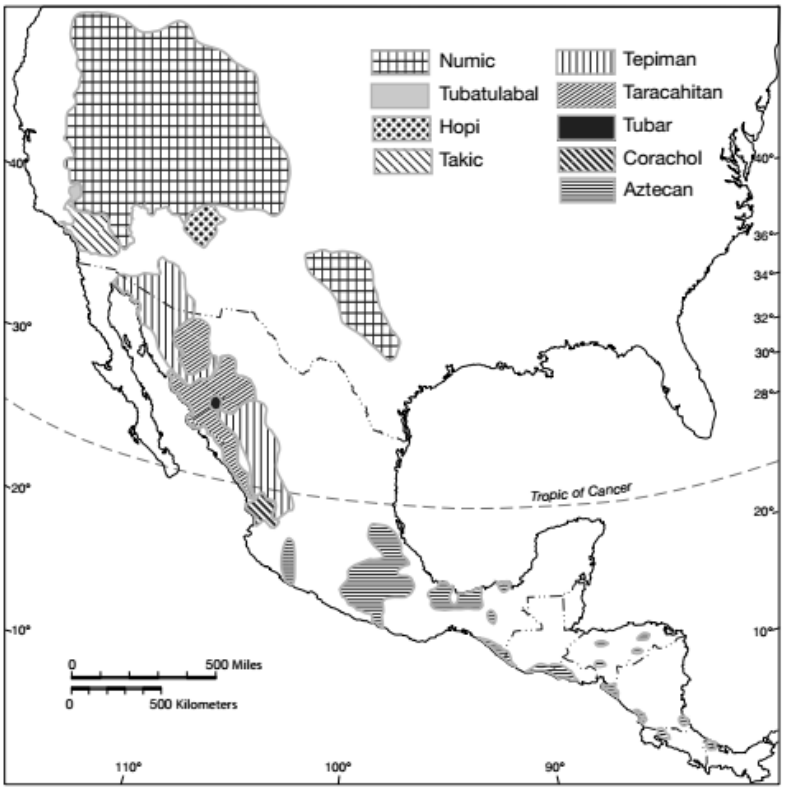


Figure 1: Map of the Uto-Aztecan Languages (Merrill 2013)

Northern Uto-Aztecan (?)		**Southern Uto-Aztecan** (?)	
Numic		**Tepiman***	
Mn	Mono	Pg	Tohono O'odham (Papago)
Pn	Tümpisa (Panamint)	Nv	Nevome
Sh	Shoshone (Gosiute)	NT	Northern Tepehuan
Cm	Comanche		
Ka	Kawaiisu	**Taracahitan***(?)	
Ch	Chemehuevi	Gu	Guarijío
SP	Southern Paiute	Tr	Tarahumara
Ute	Ute	Op	Ópata
		Eu	Eudeve
		My	Mayo
Hp	**Hopi**	Yq	Yaqui
Tbl	**Tübatulabal**	Tbr	**Tubar***
Takic (?)		**Corachol***	
Gb	Gabrielino	Hch	Huichol
Sr	Serrano	Cr	Cora
Ca	Cahuilla		
Cp	Cupeño	**Aztecan**	
Ls	Luiseño	CAz	Classical Aztec (Nahuatl)
		Te	Tetelcingo Nahuatl
		Pp	Pipil
			* = Miller's (1984) Sonoran

Figure 2. Uto-Aztecan Languages Used in this Study

3 New Data

Our work here uses K. Hill (2014)'s UA cognate sets. Expanding on Miller (1967, 1988), with input from Jane Hill and Brian Stubbs, K. Hill (2014)'s database is an extensive TshwaneLex data file arranging UA vocabulary into sets connecting presumably related UA wordforms with their meanings. The database includes 1,471 total word meanings/glosses ('nose', 'moon', etc.) across 1,274 unique cognate sets. These sets are coded using the first letter or two (vowel or consonant+vowel) and a unique identifying number, e.g., a-01 'to laugh at' thru wu-02 'strength'. The database compiles data from across the UA family and includes significant data from all of the languages included in this study. K. Hill's organization of these cognate sets and their associated meanings are taken to be independent judgments of cognation. We regard these as the most reliable cognate sets available, due to the almost 50 years of work that have gone into their creation.

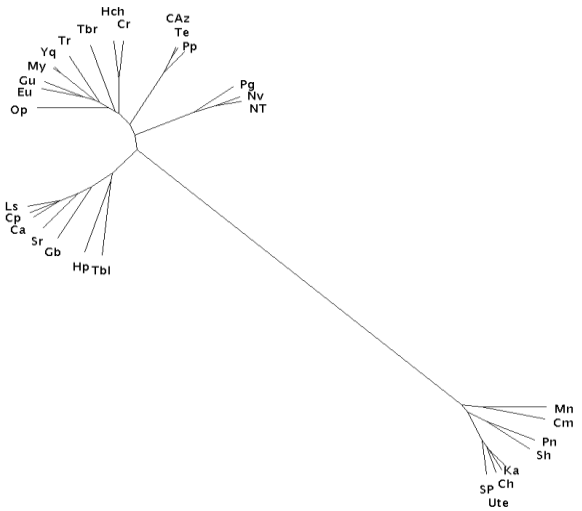
We have used the meanings in the database to relate UA cognate sets to various prominent wordlists which exist in the literature, in particular Morris Swadesh's 100 and 207 word lists, and especially Wick Miller's 100 word list, which modifies the Swadesh wordlist with some pan-UA concepts and vocabulary. Once cognate sets from the database were associated with these wordlists, we were able to run different statistical tests that have appeared in the historical linguistics literature.

We present our initial results below, introducing some traditional distance-based metrics first (§4) before turning to more contemporary Bayesian methods (§5).

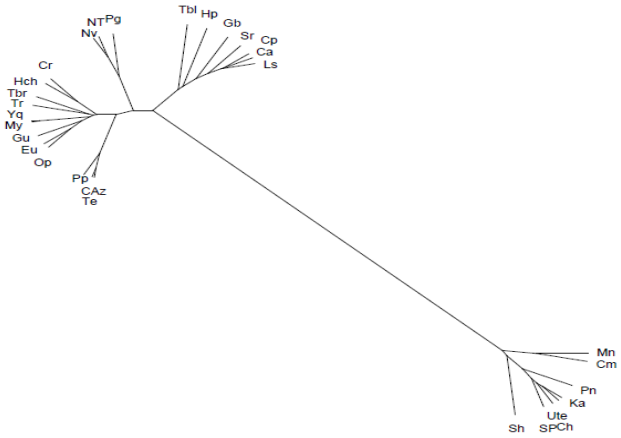
4 Distance-based Results

Miller (1984) provides possibly the simplest possible way of quantifying UA language relationships, conducting a pairwise analysis of each language in the family and counting how many words on his modified Swadesh 100 word list (hereafter, the MSw100 word list) each language had in common. We did the same procedure using the same list but then applied more sophisticated statistical methods to establish distances between languages. The two distance metrics are Euclidian (where the distance is the linear distance between 2 languages in the space) and Cityblock (where the distance is the sum of the absolute differences of each corresponding component between a two languages). The first clustering algorithm we present are Unweighted Pair-Group Method with Arithmetic Mean (UPGMA), a traditional agglomerative metric used in lexicostatistical studies (Nichols and Warnow 2008). We use the SplitsTree software to generate unrooted tree representations; having no root means that Proto Uto-Aztecan itself is not represented, but the graphs do show relative distances between individual languages and sub-groups.

(1)a. Unrooted tree derived from the UPGMA algorithm (Euclidean Distance)



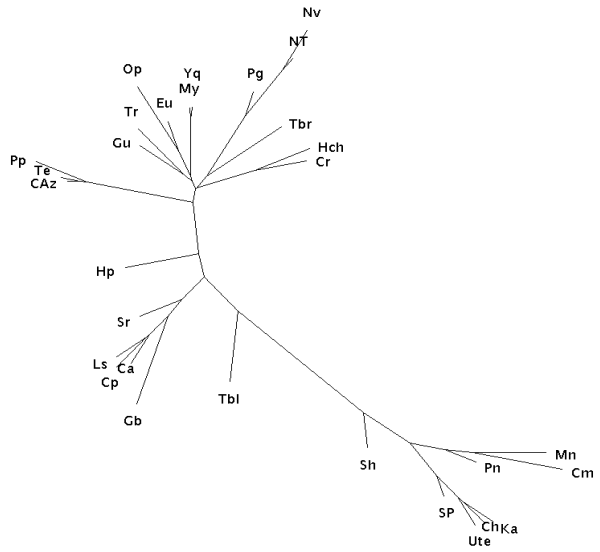
b. Unrooted tree derived from the UPGMA algorithm (Cityblock Distance)



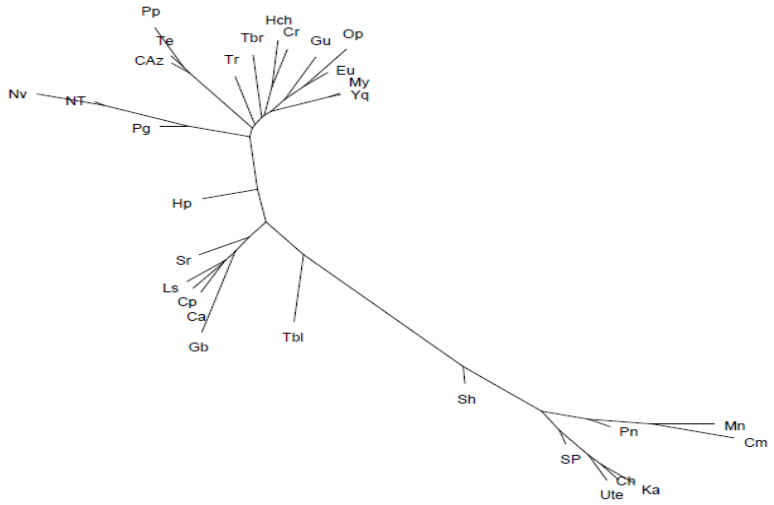
As shown in (1a) and (1b), the two distance metrics generate similar results given the UPGMA algorithm. Some of the traditional UA sub-groups are clearly identifiable using either metric, including Aztecan (Pp, Te, Caz) and Tepiman (Nv, NT, Pg). However, some groups are only found using one—cf, the conglomeration of Taracahitan languages with Corachol in (1b). Intriguingly, NUA is not found in either tree because Numic is so far away from the rest of the family. This surprising result, which seems to indicate that Numic is lexically quite different than the rest of UA, will be echoed below.

Next we present two trees using the Euclidean and Cityblock distance metrics and the Neighbor Joining (NJ) clustering algorithm. NJ (Saitou & Nei 1987) begins with a star topology and cyclically connects the nearest-most neighbor points. Connected pairs are replaced by a new point between them, which is then used for all future neighbor-joining calculations and the original points are discarded. In (2a) and (2b) we again use SplitsTree to generate unrooted trees.

(2)a. Unrooted tree derived from the NJ algorithm (Euclidean Distance)



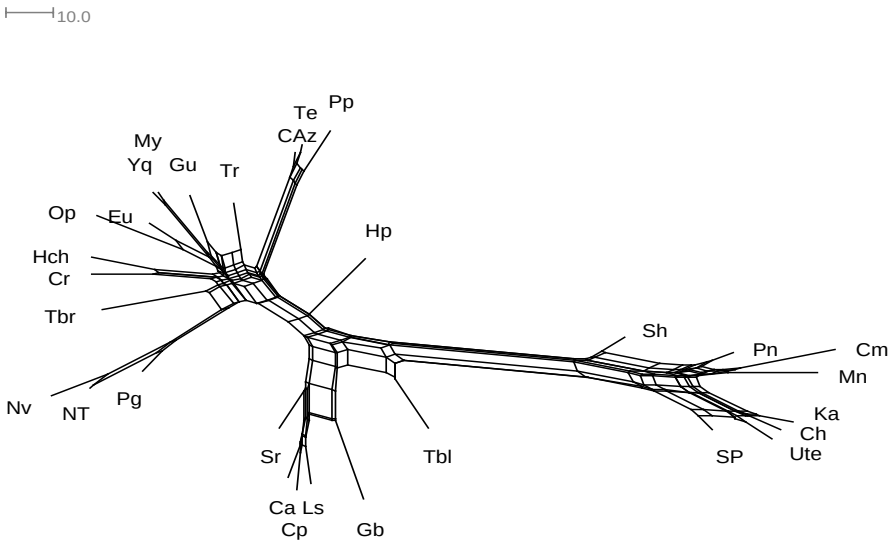
b. Unrooted tree derived from the NJ algorithm (Cityblock Distance)



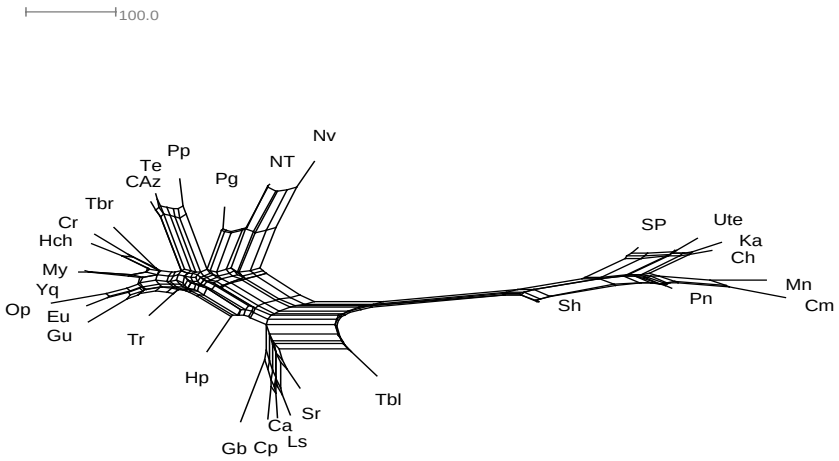
Like those in (1), these trees also seem to indicate that the Numic languages are relatively lexically distant from the rest of the family. Comparing Numic to Tübatulabal, Hopi, and Takic (i.e., Serrano, Luiseño, Cupeño, Cahuilla, Gabrielino), NUA itself seems quite diffuse in both trees compared to the more relatively closely grouped SUA languages. There does not appear to be a clear NUA cluster because the Numic languages are too distant. Likewise there is no Californian cluster, because Tübatulabal does not cluster with Takic, although Takic does appear as such. The difference between the two distance metrics is the presence of a Sonoran cluster in (2a), which is absent in (2b) because Aztecan separates Tepiman and the rest of Sonoran

Our final distance-based metric is NeighborNet (NN), which clusters similarly to NJ (beginning with star topology, cyclical pairing of closest points, unrooted) except that it allows for hypotheticals. While NJ discards the original location of two points after joining them, NN maintains those original points until they make one other connection. This allows NN to non-hierarchically represent multiple possible trees rather than forcing the results into a single tree.

(3)a. Graph derived from the NeighborNet algorithm (Euclidean Distance)



b. Graph derived from the NeighborNet algorithm (Cityblock Distance)



In (3a) and (3b) we find many of the same results as our NJ trees: Sonoran only appears in the Euclidean distance tree (3a), both trees show a SUA cluster; Tubar is either an isolate or combines with Taracahitan. This is not surprising because NN essentially shows many possible NJ trees. Perhaps the most significant differences are the presence of a Californian cluster in (3), because Hopi is separated, and in (3b) where there seem to be two main clusters of Numic vs. everything else. We find it remarkable that all of our distance-based metrics indicate that the Numic branch is by far the most divergent.

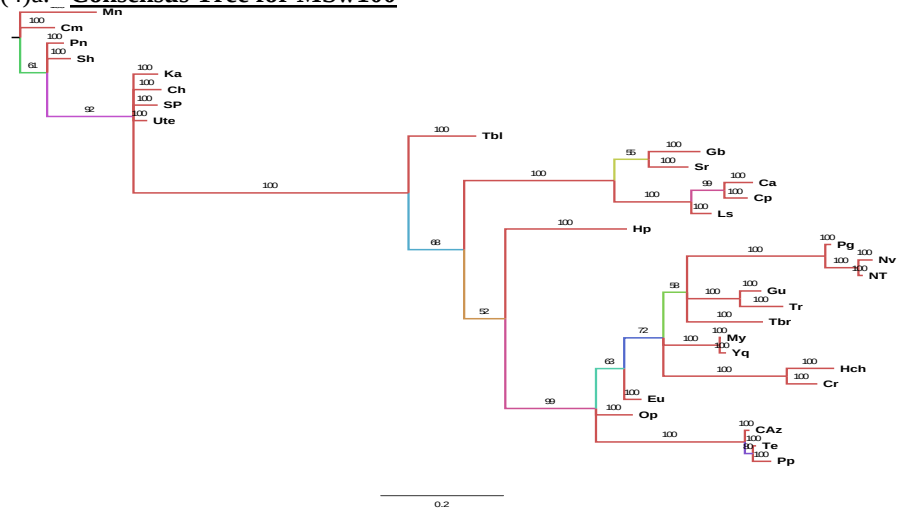
In sum, in this section we have shown the results of 2 distance metrics and 3 clustering algorithms. NUA does not appear in any of our distance-based trees, although SUA consistently does. Sonoran only appears in the Cityblock NJ and NN trees because Tepiman was otherwise separated from the other Sonoran languages. Californian does not appear as a clear branch in any of the distance-based trees. Tubar consistently groups with Taracahitan and Corachol, so that, in terms of previous proposals, these results do not support it being a Tepiman language. In the following section we present results using Bayesian character-based metrics. While in the cases above we kept the wordlist consistent (using the MSw100), in the following section we vary the list as well as the type of Bayesian method.

5 Bayesian Results

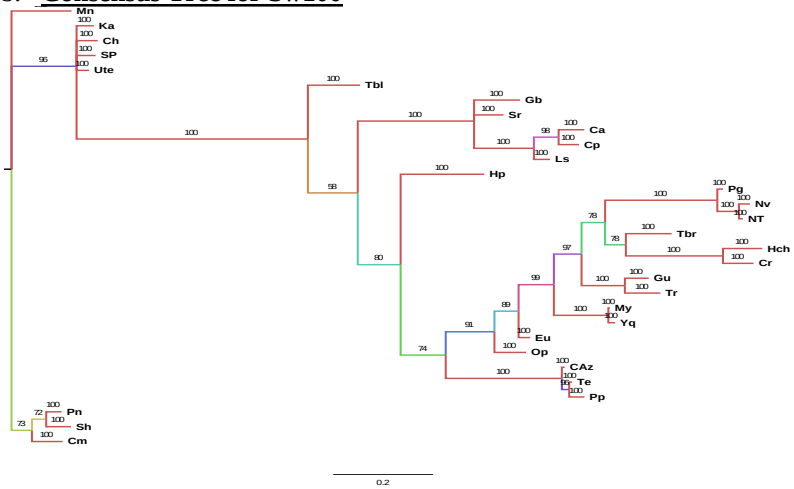
Some more contemporary approaches to quantitative historical linguistics apply Bayesian phylogenetic methods adopted from biology (Syrjänen et al. 2013, Michael et al. 2015, Birchall et al. 2016, *inter alia*). Here we present our results using Bayesian clustering across 3 wordlists: the MSw100 plus Swadesh's original 100 (1955) and 207 (1952) wordlists (Sw100 and Sw207, respectively). Perhaps of greatest surprise is that, as we found above, none of these wordlists capture Numic as a single subgroup. The Numic languages as a whole are once again shown to be lexically distant from the rest of the family.

The tree in (4c) is perhaps closest to identifying Numic, in that it groups the Western (Mono) and Central (Comanche, Shoshone and Tümpisa), however it excludes Kawaiisu from the other Southern Numic languages (Ute, Southern Paiute, Chemehuevi). The tree for the Sw100 list individually captures each of the Numic subgroups but does not group them together. Taracahitan as a subgroup is also absent from the trees in (4a-c), although the Cahitan and Tarahumara-Guarijío subgroups do appear. Californian does not appear in any of the trees because Tübatulabal is never grouped with Takic. Sonoran appears in both of the 100 lists but the Sw207 tree groups the Cahitan languages (Yaqui and Mayo) with the Aztecan languages (Classical Nahuatl, Tetelcingo Nahuatl and Pipil). As with previous lexicostatistical studies, NUA does not appear but SUA does. Finally, Tubar appears nearby Tepiman languages (Northern Tepehuan, Tohono O'odham and Nevome) and with Tarahumara-Guarijío in the MSw100 and Swadesh 207.

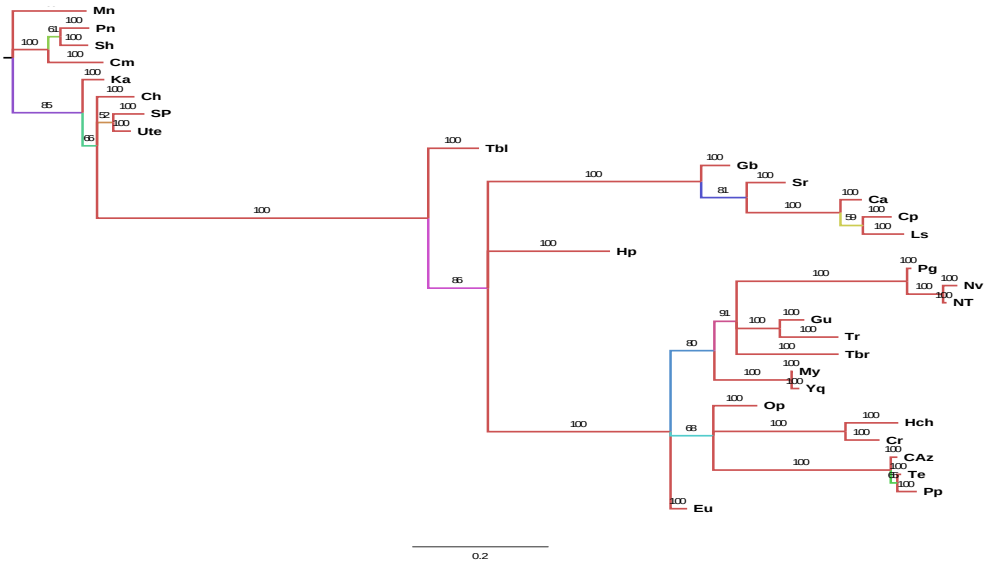
(4)a. **Consensus Tree for MSw100**



b. **Consensus Tree for Sw100**



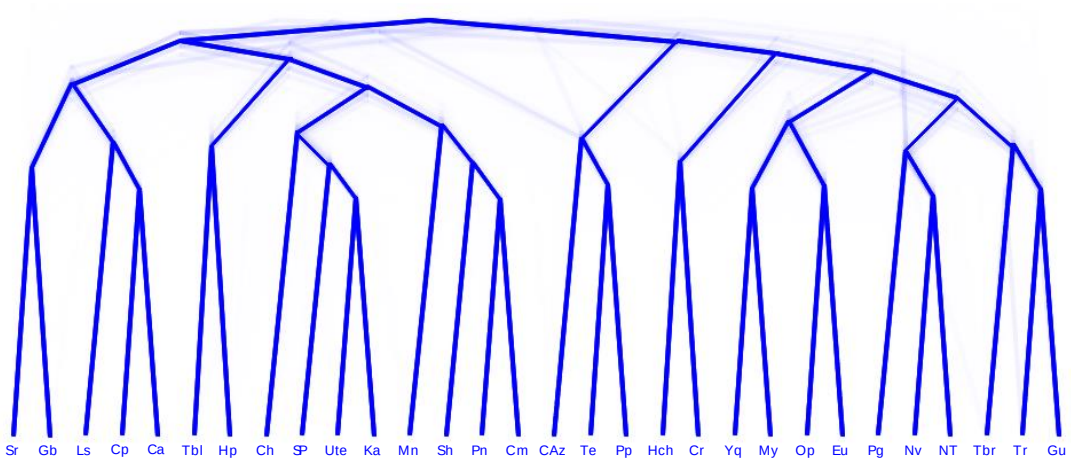
c. **Consensus Tree for Sw207**



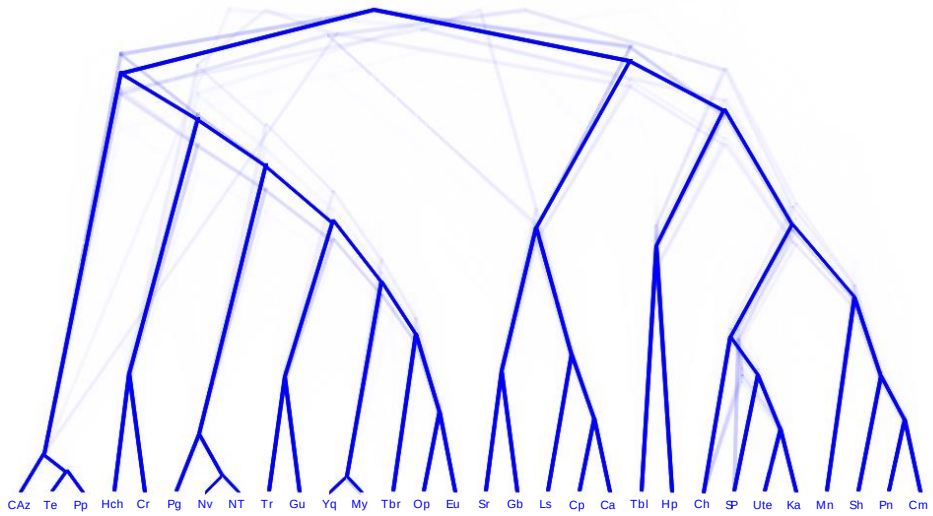
Finally, stochastic Dollo models have more recently been adopted for use in linguistic phylogenetics as a more realistic parallel to actual language change. As Bowern and Atkinson (2012) note, the stochastic Dollo model "assumes that cognates can be gained once but lost multiple times. This may be a more natural model of language change and is consistent with the assumption (upon which much of historical linguistics is based) that languages are highly unlikely to independently gain the same cognate, whereas cognates can be independently lost in descendant lineages" (pp. 828-9).

We present our results from a stochastic Dollo model of the UA languages using Sw100 and Sw207 wordlists (5a and 5b, respectively). Both trees show quite strong lines, suggesting that the identified connections are robust. NUA and SUA appear quite clearly in both trees, as does Sonoran (with Corachol at the periphery). Tubar is firmly within the Taracahitan subgroup in (5b) and groups with Tarahumara-Guarijío and Tepiman in (5a). Californian does not appear in either tree because Tübatulabal groups with Hopi.

(5)a. **Stochastic Dollo Results – Sw100**



b. **Stochastic Dollo Results – Sw207**



6 Discussion

Here we have approached the question of Uto-Aztecan subgrouping using several different phylogenetic clustering techniques—some traditional and some up-to-date—as well as three different wordlists. SUA appears in all of our trees (although it is not equally robust in each one), while NUA only appears in the stochastic Dollo trees. Tubar consistently clusters near either Tepiman or Taracahitan languages, although it only appears inside of the Taracahitan subgroup in (5b). As might be expected from Hill's (2011) critique of Taracahitan, the subgroup does not consistently appear, even within clustering algorithms. Hill's (2011) Californian subgroup also does not always appear, however, as Tübatulabal does not exclusively cluster with the Takic languages, although they somewhat cluster in the NN algorithm (3). Finally, Sonoran appears in all of the Bayesian trees (4-5) and in (2a).

In comparing our results to Miller (1984), perhaps the most remarkable finding is that our results align strikingly closely with his original findings. Miller (1984) used the measure of cognate density, a hit-count measure whereby two languages' cognate density relation was the number of cognates they shared out of the 100 word set (on the MSw100 wordlist). For clustering methods, Miller used a 40-word cutoff for subgrouping, i.e. any two languages that had at least 40 words in common out of the 100-word set were said to be a subgroup. Although we have adopted more technically advanced clustering algorithms in this paper (some of which, particularly the Bayesian methods, have been developed since Miller's original study), it is striking to us that our results nearly match Miller's (1984) main findings. Specifically, the distance-based lexical evidence analyzed here supports a grouping of SUA but not NUA, with Sonoran also appearing in most of our trees. However, NUA is identified by the stochastic Dollo method. SUA was first proposed due to lexical similarities (although see Merrill 2013 for a proposed SUA sound change), while NUA is generally supported by more traditional means (shared phonological and morphological innovations). Thus, it is possible that the SUA clade will always appear in studies privileging lexical retention.

We would like to end with a point that we made at the outset—i.e., that these findings are preliminary, and we are offering them here to further discussion of issues of UA subgrouping. We do not presume to have discovered the definitive ('God's truth') UA tree using any of the quantitative methods adopted above. But, to the extent that different methods converge on similar results we think that some of the results may be suggestive, and it is worth pondering why the results are what they are. Similarly, it is worth pondering why some expected branches, such as NUA and Numic, do not robustly appear and what processes of change might have led to such a disagreement amongst methods. We hope to continue working with K. Hill's (2014) data set, as we have only scratched the tip of the iceberg. For example, there are many larger lexical comparison sets that could be explored beyond the MSw100, Sw100, and Sw207, and it may be worthwhile to pursue different clustering techniques and algorithms. We hope that this short report supports the conclusion that the perennial questions of UA-internal subgrouping remain a fertile ground for inquiry and exploration.

References

Birchall, Joshua, Michael Dunn, and Simon J. Greenhill. 2016. A combined comparative and phylogenetic analysis of the Chapacuran language family. *International Journal of American Linguistics* 82.3: 255-284.

Bowern, Claire, and Quentin Atkinson. 2012. Computational phylogenetics and the internal structure of Pama-Nyungan. *Language* 88.4: 817-45.

Cortina-Borja, Mario, and Leopoldo Valiñas-Coalla. 1989. Some remarks on Uto-Aztecan classification. *International Journal of American Linguistics* 55: 214-239.

Cortina-Borja, Mario, Jane Stuart-Smith, and Leopoldo Valiñas-Coalla. 2002. Multivariate classification methods for lexical and phonological dissimilarities and their application to the Uto-Aztecan family. *Journal of Quantitative Linguistics* 9.2: 97-124.

Haugen, Jason D. *forthcoming*. Uto-Aztecan. To appear in *The Languages and Linguistics of Middle and Central America: A Comprehensive Guide*, ed. by Søren Wichmann. To be published in the *World of Linguistics* series, Mouton de Gruyter.

Haugen, Jason D., Michael Everdell, and Benjamin A. Kuperman. 2015. A new approach to Uto-Aztecan lexicostatistics. Presented at the Workshop on American Indigenous Languages (WAIL), University of California Santa Barbara, Santa Barbara, CA.

Haugen, Jason D., Michael Everdell, and Benjamin A. Kuperman. 2016. Bayesian phylogenetic classification of the Uto-Aztecan languages using lexical data. Presented at 13th Annual MLK Day Linguistics Symposium, The Ohio

State University, OH.
Heath, Jeffrey. 1977. Uto-Aztecan morphophonemics. *International Journal of American Linguistics* 43.1: 27-36.
Heath, Jeffrey. 1978. Uto-Aztecan* na-class verbs. *International Journal of American Linguistics* 44.3: 211-222.
Hill, Jane H. 2011. Subgrouping in Uto-Aztecan. *Language Dynamics and Change* 1: 241-278.
Hill, Kenneth C. 2014. *Wick Miller's Uto-Aztecan Cognate Sets (revised and expanded by Kenneth C. Hill*). Tshwanelex Data File. Tucson, AZ.
Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller, and Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica* 42.2: 331-354.
Manaster Ramer, Alexis. 1992. A Northern Uto-Aztecan sound law: *-c- → -y-. *International Journal of American Linguistics* 58.3: 251-268.
Merrill, William L. 2013. The genetic unity of Southern Uto-Aztecan. *Language Dynamics and Change* 3: 68-104.
Michael, Lev David, Natalia Chousou-Polydouri, Keith Bartolomei, Erin Donnelly, Vivian Wauters, Sérgio Meira, and Zachary O'Hagan. 2015. A Bayesian phylogenetic classification of Tupí-Guaraní. *Llames*. 15.2:193-221.
Miller, Wick R. 1967. *Uto-Aztecan Cognate Sets*. Berkeley: University of California Press. University of California Publications in Linguistics 48.
Miller, Wick R. 1984. The Classification of the Uto-Aztecan Languages based on lexical evidence. *International Journal of American Linguistics* 50.1: 1-24.
Miller, Wick R. 1988. Computerized database for Uto-Aztecan cognate sets. Unpublished monograph. Salt Lake City: University of Utah.
Miller, Wick R., James L. Tanner and Lawrence P. Foley. 1971. A lexicostatistical study of Shoshoni dialects. *Anthropological Linguistics*: 142-164.
Nichols, Johanna, and Tandy Warnow. 2008. Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2.5: 760–820
Saitou, Naruya, and Masatoshi Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees." *Molecular biology and evolution* 4.4: 406-425.
Sapir, Edward. 1913. Southern Paiute and Nahuatl, a study in Uto-Aztekan. *Journal de la Société des Américanistes* 10.2: 379-425.
Sapir, Edward. 1915. Southern Paiute and Nahuatl, a study in Uto-Aztekan, Part II. *American Anthropologist*, 17.1: 98-120.
Stubbs, Brian D. 2000. The comparative value of Tubar in Uto-Aztecan. *Uto-Aztecan: Structural, Temporal, and Geographic Perspectives: Papers in Memory of Wick R. Miller by the Friends of Uto-Aztecan*: 357.
Swadesh, Morris. 1952. Lexico-statistic dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos. *Proceedings of the American philosophical society* 96.4: 452-463.
Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistical dating. *International Journal of American Linguistics* 21: 121-127.
Syrjänen, Kaj, Terhi Honkola, Kalle Korhonen, Jyri Lehtinen, Outi Vesakoski, and Niklas Wahlberg. 2013. Shedding more light on language classification using basic vocabularies and phylogenetic methods: a case study of Uralic. *Diachronica* 30.3: 323-352.
Wheeler, Ward C., and Peter M. Whiteley. 2014. Historical linguistics as a sequence optimization problem: The evolution and biogeography of Uto-Aztecan languages. *Cladistics*: 1-13.

Comparative Takic: Absolutive Suffix Vocalism

Kenneth C. Hill[1], Jane H. Hill[2]
[1]Tucson, AZ, [2]University of Arizona

Absolutive suffixes appear on nouns in all the Takic languages. Their primary role is to mark non-possession. As a first approximation, these suffixes are from PUA[1] ***-ta***. While the diachronic developments of the consonants of the absolutive suffixes are reasonably well understood (1), the vocalism has remained a problem.

(1) Intervocalic ***t*** after ***i*** > TV ***y***, KI ***ts***, SE and Cupan č; other intervocalic ***t*** > TV ***r***, Serran[2] ***ċ***,[3] Cupan ***l***.[4] Postconsonantal ***t*** remains (usually with subsequent loss of the preceding, stem-final consonant).

The underlying final vowel of the suffix in Takic is usually ***a***, but reflexes of ***ɨ*** are also found in plurals. The Takic reflexes of ***ɨ*** are various: PTak ***ɨ*** > TV ***o***, Serran ***ɨ***, LU stressed ***o*** and unstressed ***u***, AC ***o***, CU ***ə*** and ***i***, CA ***e***. The development of ***i*** < ***ɨ*** is uncommon in CU but it figures in the morphology, especially in that of the absolutive. The vowel of the absolutive is subject to a general rule of apocope, whereby word-final short vowels are lost.

Writers seem to be largely unaware of the problem of the vocalism in the plural, a recent example being Mamet (2010), where the Luiseño absolutive suffix consonants are studied in great detail while making no mention of the different vowels found in the plural forms. Though the Uto-Aztecan plural suffix must be reconstructed as ***-mɨ***, most writers (e.g. Kroeber & Grace 1960, Elliott 1999, as well as Mamet) treat the vowel of the absolutive suffix as part of the plural ending: LU ***-am*** ~ ***-um***. Even if this makes sense as a synchronic account, it leaves unanswered the question of the origin of the different vowels involved.

Various phonological developments have obscured much of the diachronic basis for the differences in the consonants. This has led to a partial morphologization among the forms of the suffix and sometimes a given form of the suffix is associated with a noun stem in a way that is at odds with the diachronic development. As a consequence, some nouns now take a form of the suffix that cannot be accounted for by reference to synchronic phonology and it becomes necessary to recognize several noun classes according to suffix type.

The Takic reflexes of ***-ta*** appear in the singular and sometimes also in the plural, as in (2). Note that word-final short vowels apocopate in Takic. The word-internal tilde, ~, marks reduplication.

(2)			singular	plural with *-*ta*
a.	TV	woman	*tokoo–r*	*to~took–ra–m* (3.104.0335) [5]
b.	TV	jackrabbit	*ṡoʔii–t*	*ṡoʔii–ta–m* (3.104.0067)
c.	TV	hummingbird	*piino–r*	*pe~piino–ra–m* (3.105.0012)
d.	SE	cottonwood tree	*wama–t*	*wama–ta–m*
e.	SE	person	*taaq–t*	*taaq–ta–m*
f.	KI	person	*taaka–t*	*taaka–ta–m* (3.98.0288)
g.	LU	house	*kii–ča* [6]	*kii–ča–m*

[1] Abbreviations: AC = Acjachamem (Juaneño), CA = Cahuilla, CU = Cupeño, DCA = Desert Cahuilla, KI = Kitanemuk, LU = Luiseño, MCA = Mountain Cahuilla, PTak = Proto-Takic, PUA = Proto-Uto-Aztecan, SE = Serrano, TV = Tongva (Gabrielino). *Acjachemem* is pronounced [aˈxačəməm]. <cj> is apparently the heritage community's choice for spelling the sound [x].

[2] Serran = Serrano and Kitanemuk; Cupan = Luiseño, Acjachemem, Cupeño, and Cahuilla. There are two clades within Cupan: (1) Luiseño and Acjachemem, and (2) Cupeño and Cahuilla.

[3] ***ċ*** represents the apicoalveolar affricate; ***ṡ*** is the corresponding sibilant.

[4] Later splitting into *l* and *ly* in Cupeño-Cahuilla.

[5] See Harrington [microfilm] in References. In a citation like "3.104.0335", "3" is the volume, "104" is the reel, "0335" is the frame. Harrington numbers are provided for Tongva, Kitanemuk, and Acjachemem citations.

[6] Luiseño has "long absolutives" on some nouns. A satisfactory diachronic account for them remains to be found.

h.	LU	star	*ṡuʔ–la*	*ṡuʔ–la–m*
i.	CU	fly	*kuʔa–l*	*kuʔ–la–m*
j.	CU	jackrabbit	*suʔi–š*	*suʔ–ča–m*
k.	CU	lineage chief	*nə–t*	*nə~n–ta–m* [7]
l.	CA	deer	*suka–t*	*suk–ta–m*
m.	DCA	rattlesnake	*sewe–t*	*sew–ta–m*
n.	DCA	skunk	*tekwe–l*	*tekw–la–m*
o.	MCA	skunk	*teqwe–l*	*teqw–la–m*
p.	MCA	wildcat	*tuku–t*	*tuk–ta–m*

A second form of the absolutive suffix in Takic is from **-ti*. This form contributes allomorphs that are encountered only in the plural. The plurals of (3) show reflexes of **-ti*. Some stems are attested with both kinds of plural, as with TV 'woman' ((2a) and (3a)) and LU 'house' ((2g) and (3g)). Cupeño forms show *i* (in (3i–l)), not *ə*, the usual reflex of **i*, but the distinction between original **-ta* and **-ti* is maintained. (**-***we***-** in (3n) is the augmentative suffix; (3n) is the augmentative of (3q).)

(3)			singular	plural with **-ti*
a.	TV	woman	*tokoo–r*	*to~tooko–ro–m* (3.105.0444)
b.	LU	ant	*aana–t*	*an–tu–m*
c.	LU	jackrabbit	*ṡuʔí–š*	*ṡuʔii–ču–m*
d.	LU	fly	*kuʔaa–l*	*kuʔaa–lu–m*
e.	LU	woman	*ṡuŋaa–l*	*ṡu~ṡŋa–lu–m*
f.	LU	stone	*too–ta*	*too–tu–m*
g.	LU	house	*kii–ča*	*kii–ču–m*
h.	LU	arrow	*huu–la*	*huu–lu–m*
i.	CU	gopher	*məə–t*	*məə–ti–m*
j.	CU	dog	*awá–l*	*awá–li–m*
k.	CU	pot, dish	*kavaʔma–l*	*kavaʔma–li–m ~ ka~kvaʔma–li–m*
l.	CU	fox	*kawísi–š*	*kawisi–či–m*
m.	CA	flea	*muka–š*	*muka–če–m*
o.	CA	mountain lion	*tuk–we–t*	*tuk–we–te–m*
p.	CA	mountain sheep	*paʔa–t*	*paʔ–te–m*
q.	CA	jackrabbit	*suʔi–š*	*suʔ–če–m*
r.	CA	wildcat	*tuku–t*	*tuk–te–m*

[7] Corrected from *nəntim* in Hill and Nolasquez (1973:165 [2005:251).

The Serran examples in (2) are unusual in their retention of the absolutive before the plural suffix. The examples in (4) illustrate the more general pattern whereby the absolutive suffix is replaced by the plural suffix ***-m***.

(4)			singular	plural	PTak root
a.	SE/KI	wildcat	*tuku–t*	*tuku–m* (3.99.0358)	**tukuC*
b.	SE/KI	jackrabbit	*hwiiʔ–t*	*hwiiʔ–m* (3.98.0104)	**suuʔiC*
c.	SE/KI	bee, yellowjacket	*haaŋa–ċ*	*haaŋa–m* (3.98.0134)	**saaŋaa*
d.	SE	star	*huuʔ–ċ*	*huuʔ–m*	**suuʔu*
e.	KI	star	*huuʔ–ċ*	*huh~huʔ–m* (3.98.0061)	**suuʔu*
f.	SE	boy	*čičin–t*	*čičina–m*	**čičinaC*
g.	KI	boy	*titini–t*	*titini–m* (3.98.0087)	**čičinaC*

Examples (4b,d,e) show glottal stop metathesis. In (4d,e) this process has obscured the underlying vowel-final nature of the affected root. Example (4g) shows diminutive sound symbolism, whereby **č* > *t* and **a* > *i*.

The Serrano example (5a) shows the result of syncope with consequent cluster simplification (underlying *taŋata-t* > *taŋat-t* > *taŋat*). This results in an absolutive form that is homophonous with the possessed form (5c). The Kitanemuk form for 'sack' (5b) shows no syncope and has a reduplicated plural.

(5)			singular	plural	PTak root
a.	SE	sack	*taŋa(t)–t*	*taŋata–m*	**taŋataC*
b.	KI	sack, box	*taŋat–t*	*tah~taŋata–t* (3.98.0380)	
c.	SE/KI	my sack	*ni–taŋat* (3.98.0380)		

An exception to this cluster reduction is provided by SE *qoˑpoˑtt* 'turtle',[8] in (6), with a sequence of two released *t*'s in word-final position. This occurred once in a recording; subsequently the speaker would give only unsyncopated *qoˑpo^{R}tat*.

(6)						
a.	SE	turtle	*qoˑpoˑt(a)–t*	*qoˑpoˑta–m*	**kopotaC*
b.	KI	turtle	*kopota–t*	*kopota–m* (3.98.0117)	

In the non-Serran Takic languages, the selection of the plural form, whether representing **-ta* or **-ti*, is a lexical property of the noun. In Acjachemem, the plural form is always ***-ta-m***, ***-ča-m***, ***-la-m***, but unstressed *a* is a secondary development from **o* (< **i*) in unstressed position, and thus has merged in that position with **a*.

In Cupeño (Hill 2005:34ff) and Cahuilla, especially Mountain Cahuilla (Sauvel & Munro 1981:35–36), the choice of ending has been restructured such that the forms of the plural are largely predictable phonologically. If the vowel preceding the absolutive suffix deletes (via syncope), then the plural is in *a* (1h–n); if the vowel remains, then the plural is in **i* (CU *i*, CA *e*) (2i–n). This generalization covers the plurals that have *a* except for a few irregular forms, as seen in (7). Examples (7a,b) show irregular stems with phonologically exceptional *a* in the plural, though both have also been recorded with *i*. The first plural of (7c) is exceptional in having *a*, but it is a contraction of the second plural, for which *a* is expected. The plural of example (4d) shows the expected ending in *i* though there is irregularity in the stem. (7e) shows no exceptionality. (***-ma*** in (7a,b) is the diminutive suffix.)

(7)	CU	singular		plural
a.	boy	*kii–ma–l*		*ki~ki–ta–m*
b.	girl	*nawíš–ma–l*		*niš–ma–la–m* (but also *niš–ma–li–m*)
c.	woman	*nawíkə–t*	(1)	*na~nwi–ta–m* (but also *na~nwi–ti–m*)

[8] Serrano has three rhotic vowels: *oˑ* is low rounded and is the reflex of PUA **o*, *aˑ* is low unrounded, and *iˑ* is high unrounded. The etymological origin of *aˑ* and *iˑ* remains unclear though in some instances the rhoticity of these vowels reflects the one-time presence of **o* in a neighboring syllable.

			(2)	
d.	man	*naxáni–š*		*na~nwik–ta–m*
e.	old woman	*nišlyuvə–l*		*ni~nišlyuvə–li–m*

Other Cupeño exceptions frequently encountered are derivations in *-wənə-t*, pl. *-wən-ti-m*, as in *xwaya-xwənə-t* 'white one', pl. *xwayaxwen-ti-m*.

A few exceptions remain in Mountain Cahuilla, such as (2p) *suʔi-š* 'jackrabbit, pl. *suʔ-če-m*, but many more remain in Desert Cahuilla. *Tuku-t* 'wildcat' has the "regular" plural *tuk-ta-m* in Mountain Cahuilla (1o) while in Desert Cahuilla it has the "irregular" plural *tuk-te-m* (2q).

In the Cupeño plurals of (2) and (7d,e), instead of expected *ə*, the regular reflex of **i*, the vowel that appears is raised and fronted to *i*. This could possibly be accounted for as part of a chain shift within a morphologically restricted environment: unstressed **a* > [ə], **i* > *i*. Synchronically, however, the *i* here (and the *a* seen in other combinations) may be accounted for by vowel epenthesis, as presented in Hill (2005:32f).

In many Uto-Aztecan languages, reflexes of PUA **-ta* appear as case markers. For instance, in Cahita *-ta* is restricted to a case-marking role (Collard & Collard 1962:201–202). Among the Northern Uto-Aztecan languages, Hopi offers a more complex situation which hints at what Takic may have evolved from.

To understand **-ti*, we turn to the Hopi evidence. Hopi has two plural suffixes for nouns, *-m(i-)* and *-t(i-)*. *-m(i-)* is the only plural marker used with possessed nouns. Among non-possessed nouns *-m(i-)* marks an animate plural. *-t(i-)* is used for both animates and inanimates, but, like the accusative case suffix from **-ta*, it is used only for non-possessed nouns. The plural suffixes, besides their use in the plural, are also used in the formation of the dual. Consequently, within Hopi grammar, it is better to refer to these endings as "non-singular" (NSG) suffixes instead of as "plural" suffixes.

The dual may be an internal development in Hopi but it informs our understanding of **-ti* in Takic. The Hopi dual form normally consists of the simple form of a noun, i.e., the form used in the singular, plus a non-singular suffix, most commonly, the suffix *-t(i-)*. The plural form, when it is differentiated from the dual, entails some modification of the noun stem, usually reduplication, and animate plurals take a non-singular suffix as well, as illustrated with *taaqa* 'man' in (8).[9] (Note that many nouns, not included here, take *-m(i)* rather than *-t(i)* in the plural; some take *-t(i)* in the dual and *-m(i)* in the plural, and some take *-m* in both the dual and the plural.)

(8) Hopi nominative

singular	dual	plural
taaqa	*taaqa–t(i)*	*tàa~taq–t(i)*
man	man–NSG	PL~man–NSG

Hopi also has two accusative case suffixes, *-y(i-)* and *-t(a-)*, mentioned earlier. The accusative case forms of the noun in (8) are given in (9).

(9) Hopi accusative

singular	dual	plural
taaqa–t(a)	*taaqa–ti–y(i)*	*tàa~taq–ti–y(i)*
man–ACC	man–NSG–ACC	PL~man–NSG–ACC

The nominative dual and the accusative singular of 'man' are both *taaqat*, though in combination or in "pausalization" the difference is revealed: *taaqati* vs. *taaqata*. (The pausal form results from a special gram-matical process that preserves and often elaborates normally-deleted word-final short vowels [Whorf 1946:165]; i.e., pausalization may suspend apocope, often with added *-ʔV*.)

We suggest that a similar homophony or near homophony of inflected forms has led to the development such that **-ti* could be understood to be part of the absolutive suffix system in Takic. A plural

[9] The Hopi spellings are normalized with *i* and *ʔ* instead of orthographic <u> and <'>. Grave accent marks falling tone.

form in **-t(i)* sounds dangerously similar to a singular form in **-t(a)*. To clarify the situation, the plural form was elaborated by the addition of another plural suffix, **-mi*, assuring that there was no confusion. Such double marking of the plural category is not uncommon in Uto-Aztecan languages.

It is worth noting that in Nahuatl the plural suffix **-*tin***, illustrated in (10), is also from **-ti-mi* and is also restricted to non-possessed nouns. (Nahuatl forms are given in the traditional spelling.)

(10) Nahuatl plurals of the second declension (Carochi 1645:4v–5 [2001:33])

		singular	plural
a.	hen	*tōtol–in*	*tōtol–tin* (or *tōtol–mê*)
b.	student	*tlamachtīl–li*	*tlamachtīl–tin* (or *tlamachtīl–mê*)
c.	hunchback	*tepotzò–tli*	*tepotzò–tin* (or *tepotzò–mê*)
d.	male person	*oquich–tli*	*oquich–tin* (or *oquich–mê*)
e.	hare [jackrabbit]	*cì–tli*	*cì–tin* or *cī~cì–tin*
f.	nobleman	*pil–li*	*pī~pil–tin*
g.	lion [mountain lion]	*miz–tli*	*mī~miz–tin*
h.	rabbit	*tōch–tli*	*tō~tōch–tin*
i.	wolf	*cuetlāch–tli*	*cuē~cuetlāch–tin*
j.	youth	*tēl–pōch–tli*	*tēl–pō~pōch–tin*

Huichol *-te* ‘plural’ and Cora *-t̬e* ‘plural’ are similarly restricted (Casad 1984:233–4). A more detailed exploration of this link with non-possessed number marking would be useful in clarifying the history of Uto-Aztecan absolutives. It is clear however, that complexities in noun plural formation are a recurrent typological feature within Uto-Aztecan. Rather than wondering why most plurals show reflexes of **-ti* ‘plural’, the real mystery is why some of the plurals show reflexes of **-ta* ‘absolutive singular’.

A possible alternative synchronic analysis could hold that the variable vowel before the plural ending should instead be assigned to the plural ending itself, with the absolutive suffixes as consisting only of their consonant. This is the option chosen by Anderton (1988:81), Kroeber and Grace (1960:87), and Seiler (1977:78). But this just transfers the vowel selection problem from the absolutive suffixes to the plural suffixes. It also creates the problem of how to distinguish the behavior of the absolutive suffixes from genuinely con-sonant-final forms that take no absolutive suffix, the zero class. In Cahuilla, Seiler’s plural suffix allomorph **-*em***, does allow a seemingly straightforward treatment for consonant-final Spanish loans, e.g. *melóon* ‘cante-loupe’, pl. *melóon-em*, *kalaváas* ‘pumpkin’, pl. *kalaváas-em* (< Sp. *melón, calabaza*) (Seiler 1977:79). The **-*e***- of CA *melóon-e-m*, *kalaváas-e-m*, as well as the **-*i***- found in zero class Cupeño forms such as CU *kaxóon* ‘box’, pl. *kaxóon-i-m* (< Sp. *cajón*), however, probably should be understood within the larger Takic frame-work as a stem augment (AUG), a phenomenon similar to SE **-*ia***-, KI **-*ya***-, characteristic of plurals formed from zero-class nouns, as exemplified in (11).

(11)

			singular	plural	
a.	coyote	SE	*wahiʔ*	*wahiʔ–ia–m*	
b.	coyote	KI	*wahiʔ*	*wahiʔ–ya–m* (3.98.0096)	
c.	lamb, sheep	SE	*vareewaʔ*	*vareewaʔ–ia–m*	< Sp. *borrego*
d.	lamb, sheep	KI	*vureewɨʔ*	*vureewɨʔ–ya–m* (3.99.0367)	
e.	donkey	SE	*vuurus*	*vuurus–ia–m*	< Sp. *burros*
f.	liar	SE	*poˈqoˈavis*	*poˈh~pqoˈavis–ia–m*	
g.	sealion	KI	*kuʔmuṡ*	*kuʔmuṡ–ya–m* (3.98.0101)	
h.	Kitanemuk	KI	*akikitam* [10]	*akikitam–ya–m* (3.98.0065)	

Consideration of the zero class nouns brings us to the fact that **-*ta*** survives in Takic, or more precisely, in Serran, as a case marker, where it is used in combination with **-*yi*** to mark accusative case for nouns that take no absolutive suffix. The combination **-*ta-yi*** usually contracts to **-*ti*** in Serrano (12a) but it

[10] KI *akikitam* ‘Kitanemuk’ is remarkably similar to CU *kikitam* ‘boys’, the irregular plural of *kiimal* (6a). The KI prefix *a-* marks third person singular possessor.

retains the vowel *a* in Kitanemuk (12b). The double marking of case in ***-ta-yɨ*** parallels the double marking of plural in ***-tɨ-mɨ***, discussed above.

(12) a. SE *Iip qai yɨiŋɨʔk=kwɨnɨ* *maamċ* *wahiʔ–ti.*
here not be.much.later=QUOT\3PL>3SG hear coyote–ACC
'Here not much later they heard Coyote.'

b. KI *Wahiʔ–tay* *a–woohɨk* *kutsiʔ–t.*
coyote–ACC 3SG–bark dog–GEN
'The dog is barking at the coyote.' (3.100.0700)

Postposed Kitanemuk subject nouns appear in the genitive case as in (12b), apparently governed the personal prefix on the verb. KI *kutsiʔ* 'dog' is a zero class noun, as is SE *kučiʔ*. Asymetrically, postposed possessor nouns are unmarked: *a-mukpiʔ wahiʔ* 'coyote's nose' (3.100.0067); compare the genitive form *wahiʔ-t* in normal order *wahiʔ-t a-tuhtuaʔ* 'coyote['s] dance' (3.99.0652) (more on the Serran genitive below).

The Serran accusative from ***-ta-yɨ*** is also found with a number of possessed nouns, further demonstrating the survival of ***-ta*** as a case marker.[11] Typical examples of possessed nouns with ***-ta-yɨ*** are given in (13). Serran plurals in ***-mɨ*** are only rarely marked for accusative.

(13)

		singular	plural	accusative singular	
a.	SE	*mɨ–čuuriʔ*	*mɨ–čuuriʔ–ia–m*	*mɨ–čuuriʔ–ti*	
	KI	*mɨ–tsuuriʔ*	*mɨ–tsuuriʔ–ya–m*	*mɨ–tsuuriʔ–tay* (3.98.0365)	
		2SG–MoMo	2SG–MoMo–AUG–PL	2SG–MoMo–ACC	
		'your maternal grandmother(s), your female maternal grandrelative(s)'			
b.	SE	*nɨ–havɨʔ*	*nɨ–havɨ–m*	*nɨ–hav–ċi*	
		1SG–blanket	1SG–blanket–PL	1SG–blanket–ACC	
	KI	*ni–havɨʔ*	*mɨ–ha~havɨʔɨ*	*ni–havɨ–ċay* (pl. *ni–ha~havɨʔ–ċay* 'my clothes')	
		(3.98.0055)	(3.98.0094)	(3.99.0437)	(3.98.0094)
		1SG–blanket	2SG–PL~blanket [12]	1SG–blanket–ACC	(1SG–PL~blanket–ACC)
		'my/your blanket, clothing' (cf. *havɨ–t* 'blanket, article of clothing')			

Other Serran possessed nouns show no ***-ta*** in the accusative, only ***-yɨ***, as seen in the Serrano examples in (14). The stem variants found before suffixes in (14c–h) are not part of the inflectional process.

(14)

	SE		nominative	accusative	nominative plural
a.		my heart	*nɨ–huun*	*nɨ–huun–i*	—
b.		my fur, body hair	*nɨ–poˀh*	*nɨ–poˀh–i*	—
c.		my hand, forearm	*nɨ–ma*	*nɨ–maa–i*	—
d.		my mother	*ni–yɨʔ*	*ni–yɨk–i*	*ni–yɨkɨ–m*
e.		my father	*ni–naʔ*	*ni–naʔn–i*	*ni–naʔna–m*
f.		my older brother	*nɨ–paar*	*nɨ–paah–i*	*ni–paaha–m*
g.		my older sister	*nɨ–qooˀr*	*nɨ–qooˀh–i*	*ni–qooˀha–m*
h.		my animal	*nɨʔ–aačiʔ*	*niʔ–aašt–i*	*niʔ–aašta–m*

The Serran genitive case provides the other instance of the survival of ***-ta*** in a case-marking function within Takic. In most combinations the genitive is underlyingly the suffix ***-ɨ***. With nouns that take

[11] Anderton (1988:177) refers to the *-t-* component of this suffix complex as the "constituency absolutive". She cites an unpublished manuscript by Donald Crook, who called it a "constituency suffix".

[12] KI 'my clothes' is not attested in the nominative.

absolutive suffixes, the genitive is superficially just like the nominative because the genitive case vowel suffix **-*i*** is subject to word-final short vowel deletion, as in (15).

(15) SE *tuku–t–Ø* *a–waċ* KI *kopota–t–Ø* *a–ki*
wildcat–ABS–GEN 3SG–claw turtle–ABS–GEN 3SG–house
'wildcat claws' 'turtle's house' (3.98.0145)

In Serrano, sometimes the genitive **-*i*** occurs undeleted with a reinforcement from a following glottal stop, as in (16). This glottal stop is of unknown origin.

(16) SE *nɨiʻh–t–iʔ* *a–yiʔ* *a–maiha–m–iʔ* *pi–mi–ḱaʔ*[13]
woman–ABS–GEN 3SG–mother 3SG–child–PL–GEN 3–PL–DAT
'the woman's mother' 'to her children'

Genitives of zero-class nouns and possessed nouns that take consonantal forms of the accusative suffix, such as those of (13) above, also take a consonantal form of the genitive suffix, *–t(i), –ċ(i), –č(i)*, as in (17).

(17) SE nominative accusative genitive
coyote *wahiʔ* *wahiʔ–ti* *wahiʔ–t*
his wife *a–hiintuaʔ* *a–hiintuaʔ–ċi* *a–hiintuaʔ–ċ*
older one *atučɨniʔ* *atučɨni–či* *atučɨni–č*

These genitive forms represent another instance of the PUA case marker **–ta* but with added *–i*, though the genitive element *–i* itself is of unknown origin.

The genitive forms of most forms that take the non-consonantal accusative equally well take a non-consonantal form of the genitive, sometimes resulting in forms like those of (18) where the genitive is overtly marked only by the fact that the oblique form of the stem is used.

(18) SE nominative accusative genitive
my mother *ni–yɨʔ* *ni–yɨk–i* *ni–yɨk*
my father *ni–naʔ* *ni–naʔn–i* *ni–naʔn*

An important paradigm exists for the genitive case of the demonstratives (19), whose genitive singular is marked by the suffix SE **-*č(i)***, KI **-*ts(i)***. Unlike other consonantal genitives in **-*č*** or **-*ts***, this suffix is not phonologically motivated by an underlying *i*. Neither does this suffix correspond to consonantal accusative as with items discussed above. Also shown in (19) is the plural, which has special root forms among the demonstratives. This paradigm is provided to emphasize the distinctness of the genitive as a Serran morpho-logical category.

(19) SE nominative accusative genitive plural
this *ivi?* *ivi* *ivič* *iim*
that *amaʔ* *amai* *amač* *aam*
who *hamiʔ* *hami* *hamič* *haim*
what *hiit* *hiiti* *hiič* *hiɲim*

The accusative forms *ivi* and *hami* result from the contraction of *ivi–i* (< *ivi–yɨ*) and *hami–i* (< *hami–yɨ*); word-final long vowels regularly shorten.

[13] *Ḱ* represents a velar stop somewhat fronter than Serrano *k*; morphophonemically it is *k* with *i* coloration. The two sounds *ḱ* and *k* are distinct only in a limited set of environments: before *a* and non-prevocalically, when there is no palatal sound preceding. *Ḱ* is phonetically quite similar to Spanish /k/ and Spanish /ka/ is borrowed as *ḱa*, as in *ḱavaayuʔ* 'horse' < Spanish *caballo*.

In summary then, **-ta* and **-tɨ* are reconstructed as non-possessed noun suffixes, **-ta* marking accusative case on singulars and **-tɨ* marking plural. In most Takic languages, **-ta* has lost its case-marking feature and has crept into plural formation, joining the non-possessed plural **-tɨ* in that role in a irregular way. In Cupan, the distinction between **-ta* and **-tɨ* has restructured such that it is largely predictable from the phonological configuration of the plural stem. In Serran, **-ta* has split, becoming (a) a suffix marking non-possessed singu-lar and (b) a case marker, in combination with **-yɨ*, for the accusative singular and with *-ɨ* for the genitive singular of nouns that take no absolutive suffix, including some possessed nouns, and the non-possessed plural suffix **-tɨ* has been lost.

References

Anderton, Alice Jeanne. 1988. *The Language of the Kitanemuks of California.* University of California, Los Angeles dissertation.

Carochi, Horacio, S.J. 1645. *Arte de la lengua mexicana con la declaracion de los adverbios della.* Mexico City: Juan Ruyz. (English translation: *Grammar of the Mexican Language*, translated and edited with commentary by James Lockhart. Stanford University Press, 2001.)

Casad, Eugene H. 1984. Cora. In *Studies in Uto-Aztecan Grammar*, Vol. 4, Southern Uto-Aztecan Grammarical Sketches, Ronald W. Langacker, ed., pp. 151-459. Summer Institute of Linguistics Publication Numer 56, Vol IV. Dallas, TX: Summer Institute of Linguistics.

Collard, Howard and Elisabeth Scott Collard. 1979. *Vocabulario mayo.* Serie de Vocabularios Indígenas Mariano Silva y Aceves, Núm. 6, Castellano-mayo, mayo-castellano. Mexico City: Instituto Lingüístico de Verano.

Elliott, Eric Bryant. 1999. *Dictionary of Rincón Luiseño.* University of California, San Diego dissertation.

Harrington, John Peabody. [microfilm] The papers of John P. Harrington in the Smithsonian Institution, 1907-1057, Volume 3.

Hill, Jane H. 2005. *A Grammar of Cupeño.* University of California Publications in Linguistics Volume 136.

Hill, Jane H. and Roscinda Nolasques. 1973. *Mulu'wetam: The First People.* [Reprinted 2005 with repagination.] Banning, CA: Malki Museum Press.

Kroeber, A.L. and George William Grace. 1960. *The Sparkman Grammer of Luiseño.* University of California Publications in Linguistics, Volume 16.

Mamet, Ingo. 2010. The Luiseño Absolutive Suffix: Diachronic Perspectives. *Anthropological Linguistics* 52.3/4, pp. 239–273.

Sauvel, Katherine Siva and Pamela Munro. 1981. *Chem'ivillu' (Let's speak Cahuilla).* Los Angeles: American Indian Studies Center, UCLA.

Seiler, Hansjakob. 1977. *Cahuilla Grammar.* Banning, CA: Malki Museum Press.

Seiler, Hansjakob and Kojiro Hioko. 1979. *Cahuilla Dictionary.* Banning, CA: Malki Museum Press.

Whorf, Benjamin Lee. 1946. The Hopi language, Toreva dialect. In *Linguistic Structures of Native America*, by Harry Hoijer et al., pp. 158-183. New York, NY: Viking Fund.

Refugee Parenting and Language: An Analysis Using the Family Stress Model

Anna Holdiman and April S. Masarik
Boise State University

1 Introduction

According to United Nations High Commissioner for Refugees, 65.3 million people are displaced persons, and only 1% will be resettled into a host country (UNHCR, 2017). Despite critics and public outcry, various countries still remain committed to refugee resettlement through various government programs and agencies. The tumultuous transition from fleeing one's native country to resettling in a new country poses an excess of stressors that affect an entire family unit. Refugee parents may face the psychological effects of torture, trauma, death of family members, culture shock, economic strains and linguistic barriers after arriving to resettlement countries. Altogether these stressors may affect the way that refugees parent their children and may cause strain within the family unit.

The FSM is a theoretical or conceptual framework that guides our literature review (see Figure 1; e.g., Masarik & Conger, 2017). The FSM outlines a hypothesized stress process and has been recently adapted to study refugee families. This adapted model describes the process of resettlement in which pre-resettlement stress paired with the process of seeking asylum generates psychological distress. This distress may subsequently strain family relationships and may lead to adverse outcomes for children. Box 1 in Figure 1 refers to stress prior to resettlement (e.g., persecution, death of family members, violence and so forth). The next box (Box 2) illustrates the potential for experiencing resettlement stress. This box encompasses stressors associated with living in refugee camps, the process of seeking asylum, unsafe living conditions, uncertainty and so forth. Stresses outlined in Boxes 1 and 2 oftentimes generate psychological distress in parents (Box 3). Witnessing persecution, violence, and enduring the transition to a host country oftentimes affects parent's mental health. Subsequently, this stress may affect the dynamics in the familial unit (Box 4). In this paper, we will be reviewing existing empirical research as it relates to parenting practices, family relations, and language using the FSM, (as outlined in Figure 1). This review will primarily focus on Box 4 (Strained Family Relationships) and the major stressors affecting family relationships.

We can further identify how the resettlement process affects parenting practices and what can be done to eliminate or reduce parenting stress or help support refugee parents upon resettlement. To focus our efforts, we searched terms surrounding refugee parenting in scholarly databases (e.g., PsycINFO). Our efforts focused on peer-reviewed articles that were published after 2003.

2 Review of the Literature

The first theme of our literature review involves parent or caregiver distress (See Box 3 in Figure 1) as a result of exposure to violence and persecution prior to resettlement (Box 1) as well as stressors that come with resettlement (e.g., acculturation, discrimination: See Box 2). It is suggested that exposure to violence and persecution (Box 1) paired with acculturation process and discrimination in resettlement countries (Box 2) can lead to anxiety and depression in parents. The following theme is derived from stresses that occur as a result of variables associated in both boxes.

Proceedings of *WECOL 2017*

2.1 *Parenting Style*

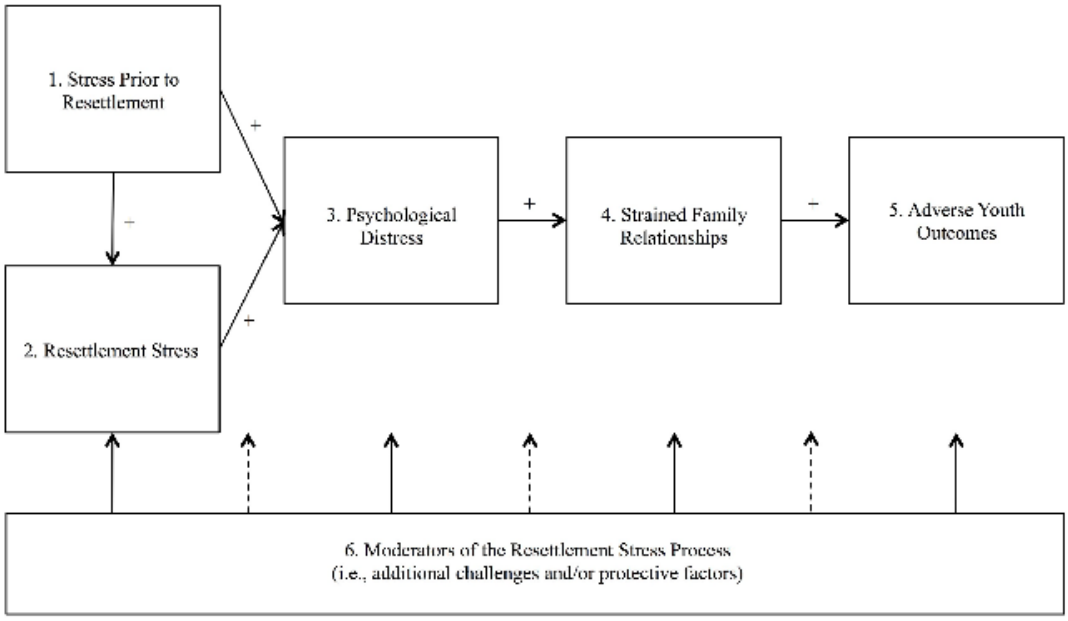


Figure 1. The family stress model of refugee family risk and resilience. This model outlines a general, hypothesized stress process by which stressors prior to and during resettlement contribute to family functioning and youth outcomes. Model adapted from Conger, Conger, & Martin (2010) and Masarik and Conger (2017).

El-Kahni, Ulph, Peters and Calam (2016) identified authoritarian (high control, low warmth) parenting practices as the most common style amongst refugee families. One explanation for this style choice is directly related to cultural values that refugees endorse. El-Kahni, Ulph, Peters and Calam (2016) conducted qualitative research on resettled Syrian families with individual interviews, followed by focus groups. Thematic analysis and the essentialist method were used for data analysis. Refugee parents identified authoritarian parenting as a means to keep children safe. Parents discussed that corporal punishment along with strict parenting practices awarded them control over their children. Some refugee parents suggested that children's play and relationships had taken a violent turn (El-Khani et al., 2016; Renzaho & Vignjevic, 2011). The violent change in children's play and relationships created fear within parents, which could suggest more control and less flexibility in parenting styles. Many parents addressed a difficulty communicating with children, further perpetuating authoritarian practices (El-Khani et al., 2016; Tingvold, Hauff, Allen & Middelthon, 2012; Santino & Marlowe, 2013).

Acculturation is the process in which one adopts behavior or cultural elements of another group. This cultural change can be a large threat to parents, because they view this change as children abandoning or losing their cultural roots (Renzaho & Vignjevic, 2011; Tingvold et al., 2012). Tingvold et al. (2012) conducted research with qualitative methods on 145 Vietnamese families. The researchers suggest that a lack of Vietnamese community support perpetuates authoritarian practices. Parent's affinity to rely solely on Vietnamese community for parenting concerns creates tension between children and parents. Moreover, this cultural reliance on resources impedes refugee's ability to take advantage of resources in their host country's culture. Additionally, Tingvold et al. suggest that conflict is due to the cultural distinctions between individualism (host country) vs. collectivism (country of origin) parenting perspectives. Refugees often flee from communities with collectivist views and resettle in countries that are primarily individualist, therefore extending conflict within the family unit (Tingvold et al., 2012).

Parents may implement stricter rules or try to control aspects of children's life such as eating habits, after school activities, music, social activities, and so forth in efforts to minimize assimilation or acculturation (Renzaho et al., 2011). Santino and Marlowe (2013) studied South Sudanese parents (four single mothers, and two married fathers) through in-depth interviews and thematically analyzed data. They

then identified a generational deficit between parents and children as a major source of conflict. Children's interest in the host culture's music, social media, and other technology is an example of a major generational difference. It is not uncommon for mainstream children in any culture to have interest that do not align with parents. This gap paired with a new culture can be a major stress because children have access to technology, music, and activities that parents have not been exposed to. This imbalance between parents and children broadens due to the plethora of resources that children are granted in a school setting that parents may not have access to (Santino & Marlowe, 2013). Furthermore, acculturation is often expedited among refugee children, due to vast exposure to the host culture on a daily basis from school related activities, social media and less ties to their ethnic background compared to their parents (Santino & Marlowe, 2013).

2.2 *Family Relations* The second theme of focus is family relations. This theme originated from Box 4 of the FSM, family distress. A study conducted by Atwell, Giggord and McDonald-Wilmsen (2009) explored the relationship between refugee parents and children focusing on parent's perceptions of their children's futures. Atwell discusses the pressures on families after resettlement due to different opportunities and support children receive. Children's schooling provides them with an array of resources (education, technology, school clubs, ESL teachers and so forth). Parents often do not have access to such resources. Additionally, the researchers discussed how strain could be placed on families due to differing parent and child expectations of what their future should hold. The study focused on the 'sense of coherence' model (SOC) to help establish the importance of appropriate expectations and visions for their children's futures. The uncertainty throughout the resettlement process can make thinking about the future daunting or even obsolete. Many families have been living on a day-to-day basis struggling to satisfy basic needs. After arriving to a host country and satisfying these basic needs (food, safety, and shelter) it can be difficult for parents to think about what the future holds for their children. The study consisted of refugee children or teens along with a family member they identified as a caregiver. The sample size consisted of ten male and female caregivers from South Sudan, Burma, Afghanistan, and Liberia. They were interviewed for three months using qualitative and quantitative visual, oral and written tools to gain demographic data as well as other psychosocial resettlement information. The study examined ways to eradicate conflict when a common vision was not shared between parents and their children. Family roles shift due to changes in traditional hierarchical systems. Traditionally, fathers are disciplinarians and mothers are nurturers or handle domestic work. Refugees fleeing war-torn countries or persecution may have lost family members, may not be able to seek refugee together, and may have members still residing in their country of origin (Atwell et al., 2009). A mother, or father may now be a single parent and struggle balancing the opposing gender role. Parents can be concerned about relatives still in danger, feel uncertainty or overwhelming unpredictability in their family's life. Refugees also recognized childrearing as a community effort. This could be difficult due to a lack of ethnic communities in host countries (Atwell et al., 2009).

Lewig, Arney and Salveron (2010) conducted a pilot study to further understand parenting in a refugee context. The sample consisted of 55 child protection practitioners, as well as 130 refugee participants from eight different ethnic communities. Participants were 17- 62 years old. Survey, interviews, and focus groups were the data collection methods used and were created by members of the South Australian statutory child protection authority. A survey by practitioners was established to identify variables that help and hinder refugee families and suggested more effective practices. Questions were open-ended and allowed practitioners to elaborate. Additionally, in-depth phone interviews were used to gain more information from practitioners.

Semi-structured focus groups lasting between 1.5- 6 hours were the methods for data collection. Participant information was discussed, translated, and analyzed (qualitatively and thematically) by a research cohort. They found that culturally relevant childcare was a barrier. Parents stayed home with children instead of attending classes, working, or gaining resources to gain independence (Lewin et al., 2010). Subsequently, one of the greatest stressors identified was the power differential that occurs between children and parents. Children oftentimes assimilate faster due to schooling and consequently have greater access to resources. They may be the primary provider for the family, serve as a translator, or parent younger siblings (Atwell et al., 2009). This change in the power structure or hierarchy often threatens parents and can cause them to adopt harsher parenting practices to gain more control.

2.3 *English Language* Another factor affecting distress in families are language barriers (Atwell et al., 2010; Lewig et al., 2010). Parents acknowledged that language served as a barrier to harnessing services, employment or resources from professionals. Parents identified language as the most difficult obstacle upon arrival to host countries. Furthermore, referring to the aforementioned power deferential, language can be a major factor in broadening the gap between parents and children post-resettlement. Lewig et al. identified language barriers between practitioners and refugees as a deterrent for taking advantage of programs or resources. Additionally, they identified that practitioners and social service providers often seemed culturally incompetent and did not understand families' frustration upon arrival. Practitioners also identified the lack of cultural competence as a hindrance in their ability to provide for families. Both acknowledged a need for better parent child communication and collaboration between families, schools, and the community. The need for culturally relevant English classes was seen as a potential solution (Lewig et al. 2010).

3 Summary

At present, we find that continued research among refugee families is warranted. The multidimensional nature of stressors associated with parenting demands further research to best help refugee families. All aspects of life may change upon refugees' arrival and therefore resettlement cannot be simply viewed from one angle. Cultural, linguistic, occupational, and familial unit change cannot be attributed to one particular aspect of resettlement, but rather the cumulative effect of change to all aspects of one's life.

Further research is also warranted in terms of the psychological wellbeing of parents. It is important to examine the family unit using a model, like the FSM, to analyze multiple variables that affect stress within the family. The mental health and wellbeing of parents directly affect the outcome of children. Mental health shapes the way in which parents financially provide and support children. This aspect of resettlement has a domino effect on children. If researchers and practitioners want to help families, they need to address the ways in which the parent-child relationship could be damaged or strengthened to promote post-resettlement success. Additionally, the implementation of longitudinal studies could be beneficial in observing if parental stress changes throughout time or the longer one resides in a host country. The creation of culturally relevant parenting classes or orientation prior to resettlement could help eliminate these issues or expose refugee families to potential resources provided to aid (Lewig et al., 2010). Furthermore, it is important to focus research efforts on protective factors that disrupt the stress cycle in refugee families (i.e., resilience). The importance of cultural education for community members (practitioners, teachers, social workers) can help those working with this population be more aware and sensitive to refugee needs. Practices could be employed that are more culturally aligned with refugee's country of origin in order to help refugees feel more comfortable and subsequently be more successful. Continued research could theoretically ignite transformation of the resettlement process and ease the often tumultuous transition (Lewig et al., 2010).

References

Atwell, R., Gifford, S. M., & McDonald-Wilmsen, B. (2009). Resettled refugee families and their children's futures: Coherence, hope and support. *Journal of Comparative Family Studies, 40*(7), 678-698.

El-Khani, A., Ulph, F., Peters, S., & Calam, R. (2016). Syria: the challenges of parenting in refugee situations of immediate displacement. *Intervention, 14*(2), 99-113.

Lewig, K., Arney, F., & Salveron, M. (2010). Challenges to parenting in a new culture: Implications for child and family welfare. *Evaluation and Program Planning, 33*, 324-332.

Masarik, A. S., & Conger, R. D. (2017). Stress and child development: A review of the Family Stress Model. *Current Opinion in Psychology*, *13*, 85-90.

Renzaho, A. M. N., & Vignjevic, S. (2011). The impact of a parenting intervention in Australia among migrants and refugees from Liberia, Sierra Leone, Congo, and Burundi: Results from the African Migrant Parenting Program. *Journal of Family Studies,17*, 71-79.

Santino Atem Deng & Jay M. Marlowe (2013). Refugee resettlement and parenting in a different context, *Journal of Immigrant & Refugee Studies, 11*, 416-430, DOI: 10.1080/15562948.2013.793441

Tingvold, L., Hauff, E., Allen, J., & Middelthon, A.-L. (2012). Seeking balance between the past and the present: Vietnamese refugee parenting practices and adolescent well-being. *International Journal of Intercultural Relations, 36*, 563-574.

United Nations High Commissioner for Refugees (n.d.) Figures at a Glance. Retrieved November 16, 2016, from http://www.unhcr.org/en-us/figures-at-a-glance.html

Zimmermann, S. (2012). Understanding repatriation: refugee perspectives on the importance of safety, reintegration, and hope. *Population, Space and Place, 18*, 45-57.

"And Lots More Beside[s]": Analysis of *besides* and *in addition to* in American and British English Corpora

Stacey M. Jacobson

The University of Alabama

1. Introduction

Many students, both domestic and international, attempt to augment their writing with the use of less frequent (though often more nuanced) linking adverbials, thus hoping to create prose that sounds more academic or educated, but often sounds awkward or non-native. An adverb that is frequently misunderstood, if not misused, is *besides*. Alongside its many meanings, *besides* often requires a different syntactic structure than *in addition to*, a transitional phrase that shares meaning with *besides*, but is typically less problematic to insert into a given sentence due to restrictions on its placement. Owing to an increase in the desire to study in English speaking countries, the need to be successful in these settings is paramount, leading students to try to distinguish themselves and demonstrate proficiency by using elevated vocabulary. For this reason, issues with these linking adverbials and transition words or phrases are often seen by teachers grading papers and writing assignments or raters scoring standardized exams with a writing component. This study aims to investigate whether the issues seen with the use of the linking adverbial *besides* correlate to usage in either American or British English.

2. Literature Review

Overuse or misuse of lexical items is a common occurrence in L2 writing, particularly at more advanced levels where vocabulary has more nuanced and complex meanings. This can be even further compounded by the lexical variation that exists between varieties of a language (Sanyal, 2013). One such lexical category that often leads to misuse by L2 writers is linking adverbials; *besides* falls into this category. Gao (2016) investigated the academic English writing of English and Chinese native speakers, focusing specifically on linking adverbials. To do so, they created a corpus comprised of academic research articles from various fields and searched for the linking adverbials therein. They found that the English and Chinese native speakers did not differ in the overall amount of linking adverbials they used, but they did find significant differences in how the additive and adversative linking adverbials were used, with *besides* appearing among those in the additive classification, which they claim reflects differences in argument style between English and Chinese scholars, with additive linking adverbials appearing more frequently in the academic writing of the English native speakers. In terms of raw frequency, *besides* appeared in their created corpus four times in the writing by English native speakers, but 83 times in the articles written by Chinese native speakers (Gao, 2016, p. 25).

Yeung (2009) also investigated the linking adverbial *besides*, looking at both meaning and use, in native English corpora and in a corpus comprised of the academic writing of Hong Kong English learners. They made use of the Cobuild Database, and editorials written in multiple varieties of English, as well as research articles and abstracts. The academic writing for the English learners was collected in the form of argumentative and expository essays. They found that English learners in Hong Kong included *besides* in their formal writing with much eight times the frequency of the number of instances of *besides* in the native English corpora. However, the function of *besides* was very limited; the Hong Kong English learners did not move beyond the meaning of *in addition* in their usage.

Additionally, Lei (2012) found that Chinese doctoral students overused many linking adverbials, and typically incorrectly used the linking adverbials *besides* and *actually* in their formal writing. They collected 20 doctoral dissertations in applied linguistics whose authors native language was Chinese, in conjunction with another 120 academic articles published in six English journals of applied linguistics. They found that linking adverbials were the most frequently used linking adverbials by both groups, though *besides* fell

outside of the ten most frequent linking adverbials in both groups, though they did list it as one that was frequently overused by the students. They also marked it as misused, attributing this to the overextension of speech register into writing, wherein *besides* appears to be quite salient, and its presence in Chinese textbooks and teaching materials.

Pertaining to textbooks and materials available and in use for foreign language instruction internationally, Leedham and Cai (2013) investigated the use of linking adverbials in writing samples from British and Chinese students at universities in the United Kingdom, utilizing the BAWE corpus which is comprised of undergraduate university students' writing. They supplemented this corpus with writing samples from Chinese students studying abroad. They found that *besides* was mainly used in order to add information, though they identified it as informal and thus unsuitable for academic writing. Attempting to explain this frequency, they analyzed the model answers from 15 papers from the NMET. They found that these sample texts from the Chinese university entrance exams (NMET) contained large numbers of instances of *besides*, which might influence the saliency of this particular linking adverbial to this group of non-native speakers.

3. Research Questions

Due to acknowledged differences between American and British English, and the use of British textbooks in English as a Foreign Language classrooms internationally, this study aims to investigate the following questions regarding the use of *besides* and *in addition to*:

1. How has the usage of *besides* and *in addition to* changed over time in American English?
2. Is there a difference in the usage of *besides* and *in addition to* in British versus American English?
3. Is there a difference between the uses of *besides* and *in addition to* in different mediums (i.e., spoken English versus academic English)? Is this usage similar or different in British and American English?

This study attempts to answer these questions from the standpoint of frequency of use; corpora were consulted and searched in order to ascertain any similarities or differences between both the terms and the varieties of English.

4. Methodology

4.1 *Corpora* Three corpora were consulted in this study: one for the historical timeline, and two to represent dialectal variation in English. The Corpus of Historical American English (COHA) was selected to shed light on the historical trend of the usage of *besides* and *in addition to* in American English. (Davies, 2010-) The COHA corpus contains 400 million words and spans three centuries, beginning in 1810 and ending in 2009.

The other two corpora were selected as representatives of modern spoken and written English in America and Britain, respectively. The Corpus of Contemporary American English (COCA) represented modern American English usage; the tokens were divided into categories: Spoken, Academic, Fiction, Magazine, and Newspaper. (Davies, 2008-) The British National Corpus (BNC) provided a picture of modern British English; this corpus divided its tokens into the following categories: Spoken, Academic, Fiction, Magazine, Non-Academic, and Miscellaneous. (Davies, 2004-) In total, COCA contains 520 million words, and BNC contains 100 million words; the breakdowns of the corpora by category can be found in Appendices A and B, respectively.

4.2 *Search Procedure* In both COCA and BNC, the tokens *besides* and *in addition to* were searched using the *chart* feature, which gave the total number of tokens in addition to splitting them up by category. This provided the raw data by section, which was necessary in order to compare the spoken and academic categories, as well as providing the totals. In COHA, the *chart* feature was also used during the search, though it showed the usage per decade as well as the total.

4.3 *Normalization and Analysis of Data* Due to the disparity between the sizes of the COCA and BNC corpora, the data from the COCA corpus had to be normalized in order to allow accurate comparisons to the data from the BNC corpus. Specifically, because the COCA corpus contains 520 million words, whereas the BNC corpus only contains 100 million, the COCA corpus was divided by 5.2 before any

comparisons were made between the two corpora. Chi-squared tests were performed on the data to determine whether there are significant differences in the uses of *besides* and *in addition to* both within and between the corpora.

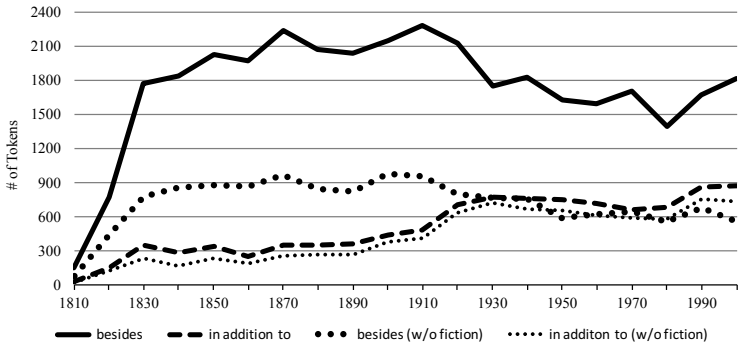
5. Findings

5.1 *Historical Trends* Historically, *besides* has been used much more frequently in American English than *in addition to*. The highest number of tokens for *besides* occurred in 1910 (N=2279), and another large spike had occurred in 1870 (N=2230). Conversely, between 1910 and the 1930s, there was a sharp decrease in the number of tokens of *besides* (N=1744), and the lowest point of decline occurred in 1980s (N=1387). However, in the 2000s it seems that *besides* is once again on the rise, the number of tokens surpassing what it was in the 1970s (2000s: N=1811).

On the other hand, *in addition to* appears to have been almost consistently increasing in use, according to the data in COHA. The first sharp spike of usage of *in addition to* occurred in the 1920s, where the number of tokens jumped from 480 (in the previous decade) to 704. The second sharp spike occurred between the 1980s and the 1990s, as the number of tokens increased from 679 to 862; between the 1990s and the 2000s, however, the number of tokens appears to have plateaued. Regardless of the increase and decline in usage of both of these terms, it is clear that *besides* has been used much more often historically than *in addition to*, and that this pattern appears to continue the disparity between the uses of these terms.

However, this analysis does not take into account the effect of genre on the frequency of *besides*. In total, there were 34,742 tokens of *besides* in COHA; when those classified as pertaining to the Fiction genre were removed, there were 14,396 tokens remaining. As can be seen in Figure 1 below, the removal of the tokens from this genre causes a remarkable shift in the patterning of *besides* historically.

Figure 1: COHA data trend, with and without Fiction category



In contrast to looking at the data as a whole, removing the tokens from the Fiction genre provides a clearer picture of the historical trend of *besides,* as well as insight into a possible explanation for the aforementioned disparity between the historical usage of *besides* and *in addition to*. It is important to note that the difference between the instances of *besides* when Fiction is included versus when it is not is highly statistically significant ($[X^2(1,N=24569)-8424.432, p<0.001]$), an indication that the use of *besides* in the Fiction genre historically may have made use of the nuanced meanings of *besides* as well as its syntactic flexibility, which could account for its prevalence, as, beyond the meaning of *in addition to*, *besides* can also be used to mean *otherwise, over and above,* and *as well* (Besides, n.d.). Fiction writing may have also favored *besides* for its potential for informality (Leedham & Cai, 2013), something that could have been especially useful in creating different registers for characters. Though its decline appears to have begun around 1910, it seems that modern fiction texts may still utilize *besides* in similar ways. Interestingly, *in addition to* was also found to be statistically significantly different when comparing the data with and without the inclusion of the Fiction tokens ($[X^2(1,N=24569)-8424.432, p<0.001]$), though the number of tokens of *in addition to* was smaller (N=10,132 with Fiction included, N=8,467 without it), indicating that *in addition to* is not absent from the Fiction genre.

5.2 *Comparisons Within COCA* A chi-squared analysis of the numbers of tokens of *besides* and *in addition to* categorized as *spoken* and *academic* in COCA found a significant difference between the use of these terms in spoken and academic American English. As shown in Table 1 below, there is an overwhelming preference for the use of *in addition to* in academic writing.

Table 1: Comparison of Spoken and Academic English in the COCA Corpus

	Spoken	Academic	Total
besides	1978 (1151.63) [592.97]	2393 (3219.37) [212.12]	4371
in addition to	2457 (3283.37) [207.98]	10005 (9178.63) [74.4]	12462
	4435	12398	16833

[X^2(1,N=16833)-1087.48, p<0.001]

As exemplified in Table 1, there is only a small difference in the usage of *besides* and *in addition to* in spoken American English, whereas there is an enormous difference in academic American English, and the difference between the usage of the two words is highly significant ($p < 0.001$). Interestingly, the quantity of tokens of *besides* in both spoken and academic American English is very similar, further demonstrating that *in addition to* is highly favored in academic settings.

5.3 *Comparisons Within BNC* The BNC corpus, upon first glance, appears to have a similar distribution of the use of *besides* and *in addition to*, albeit with far fewer tokens (due to the fact that the BNC corpus is 5.2 times smaller than the COCA corpus). However, there are some striking differences, which can be seen in Table 2 below.

Table 2: Comparison of Spoken and Academic English in the BNC Corpus

	Spoken	Academic	Total
besides	76 (54.7) [8.3]	324 (345.3) [1.31]	400
in addition to	106 (127.3) [3.5]	825 (803.7) [0.56]	931
	182	1149	1331

[X^2(1,N=1331)-13.74, p=0.00021]

As was seen in the COCA corpus earlier, there is also a significant difference between the uses of *besides* and *in addition to* in spoken and academic contexts in British English. It is interesting to note, however, that there are comparatively fewer tokens of both in spoken British English.

5.4 *Inter-Corpus Comparisons* The tokens of *besides* and *in addition to* were also compared between the COCA and BNC corpora in order to determine if there were significant differences between their usage between British and American English. Table 3 contains the comparison of the number of tokens in spoken American and spoken British English, while Table 4 contains the academic English tokens. In order to obtain a clearer picture of the differences between the two corpora, the COCA data were normalized for these two comparisons.

Table 3: Spoken American English Compared to Spoken British English

	COCA - Spoken	BNC - Spoken	Total
besides	380 (375.81) [0.05]	76 (80.19) [0.22]	456
in addition to	473 (477.19) [0.04]	106 (101.81) [0.17]	579
	853	182	1035

[X^2(1,N=1035)-0.47, p=0.49]

From the data in Table 3, it is clear that there is no significant difference ($p = 0.49$) between the use of *besides* and *in addition to* in spoken American and British English, though both are still more frequently

found in spoken American English. Conversely, there is a significant difference in usage in academic English, as can be seen in Table 4 below.

Table 4: Academic American English Compared to Academic British English

	COCA - Academic	BNC - Academic	Total
besides	460 (529.03) [9.01]	324 (254.97) [18.69]	784
in addition to	1924 (1854.97) [2.57]	825 (894.03) [5.33]	2749
	2384	1149	3533

[X^2(1,N=3533)-35.59, p<0.001]

In contrast to Table 3, the difference in the use of *besides* and *in addition to* when comparing academic British and American English is highly significant ($p < 0.001$), though both favor the use of *in addition to* in the academic setting.

6. Discussion and Conclusions

In most cases, the data point to an overwhelming preference of both *besides* and *in addition to* in American English, with *in addition to* being favored even more than *besides.* Nowhere was this clearer than in the COCA results comparing spoken and academic English; the raw number of tokens of *in addition to* was over 10,000 whereas the number of other tokens in comparison were no larger than 2,500 in any other category. However, the current usage of these forms does not match the historical trend, in which *besides* is still the overwhelming favorite, though *in addition to* has been steadily gaining ground. Modern British English, on the other hand, seems to prefer *in addition to* in most contexts, though the usage of both is not as frequent as in Modern American English.

The favoritism attributed to the use of *besides*, however, appears not to represent the entire situation. When the category of Fiction was removed from the data, *besides* patterned similarly to *in addition to* beginning around the 1920s. This suggests that exposure to fiction may be an explanation for L2 errors in the use of *besides*. The use of *besides* in modern American and British English was not statistically significant when comparing the categories Academic and Fiction ($p = 0.205$); previous research has already found that *besides* is favored in L2 academic writing (Lei, 2012; Gao, 2016), as well as in test preparation and teaching materials (Leedham & Cai 2013; Yeung 2009). Thus, it stands to reason that if *besides* is currently shown to be used with high frequency in academic writing, and this study has found no significant difference between the Academic and Fiction categories, L2 learners' exposure to fiction could have caused this favoritism and overuse, especially if there were no guidance from teachers available for the students to utilize when they encountered a word used in an unfamiliar way, or if said guidance was insufficient or flawed itself.

Upon further investigation of the usage of *besides*, it is less surprising that *besides* would be favored in fiction, as its meaning is much more nuanced and its syntactic location is more varied than *in addition to*. This syntactic flexibility is a feature that is highly desirable in fiction writing, as is the use of nuanced or semantically complex words, a category into which *besides* fits. Whereas *in addition to* can only occur at the beginning of a sentence, typically as part of a subordinate clause (or at the beginning of nonrestrictive elements), *besides* can be positioned at the beginning of a clause and offset with a comma, at the end of a clause, and offset by commas as a nonrestrictive element within the clause. Semantically, other than *in addition*, three of the meanings of *besides* are *otherwise*, *over and above*, and *as well* (Besides [Def. 2, 3, 4], n.d.). These uses of *besides* carry much semantic nuance and, as such, would be difficult for English learners to grasp, particularly without overt instruction or demonstration of the differences, thus leading to the overuse of *besides* in an attempt to mean *in addition*, as this is the most straightforward and readily understood definition. Therefore, it is plausible to posit that unfettered exposure to vocabulary through fiction has contributed to the misuse of *besides*.

References

Besides [Def. 2, 3, 4]. (n.d.). In *Merriam Webster Online,* Retrieved September 27, 2016, from http://www.oed.com/view/Entry/18071?redirectedFrom=besides#eid

Davies, M. (2004-) BYU-BNC. (Based on the British National Corpus from Oxford University Press). Available online at http://corpus.byu.edu/bnc/.

Davies, M. (2008-) The Corpus of Contemporary American English: 520 million words, 1990-present. Available online at http://corpus.byu.edu/coca/.

Davies, M. (2010-) The Corpus of Historical American English: 400 million words, 1810-2009. Available online at http://corpus.byu.edu/coha/.

Gao, Xia. (2016). A cross-disciplinary corpus-based study on English and Chinese native speakers' use of linking adverbials in academic writing. *Journal of English for Academic Purposes,24*, 14-28.

Joseph and the Amazing Technicolor Dreamcoat script by Tim Rice (n.d.). *Scripts.com.* Retrieved September 27, 2016, from http://www.scripts.comscript/126.

Leedham, Maria, & Cai, Guozhi. (2013). Besides ... on the other hand: Using a corpus approach to explore the Influence of teaching materials on Chinese students' use of linking adverbials. *Journal Of Second Language Writing*, *22,* 374-389.

Lei, Lei. (2012). Linking Adverbials in Academic Writing on Applied Linguistics by Chinese Doctoral Students. *Journal of English For Academic Purposes*, *11*(3), 267-275.

Sanyal, D. (2013). A Comparative Analysis of Lexical Variation in American and British English with special reference to few selected words. *Rupkatha Journal On Interdisciplinary Studies In Humanities*, *5*(3), 177-188.

Yeung, Lorrita. (2009). Use and misuse of 'besides': A corpus study comparing native speakers' and learners' English. *System*, *37,* 330-342.

Karuk Agreement

Elango Kumaran
UC Berkeley

Karuk's verbal agreement morphology, with its non-canonical dependence on subject and object, mood, and polarity, offers unique insight into the grammatical processes underlying agreement. Karuk has received some attention from agreement theorists (Béjar 2003, Campbell 2012)[1], but until now no one has formulated an explicit derivational account of the entire system. This paper does that.

Béjar accounted for a small subset of the data using a cyclic Agree framework; equipped with a more complicated theory of cyclic Agree, Campbell was able to account for a medium-sized subset. In this paper, I clarify and refine Campbell's analysis, and formulate an extension of Campbell's analysis to cover the entire Karuk agreement system; and then I argue in favor of a different analytic approach, without cyclic Agree.

1 Introduction

Karuk[2] is an isolate from northern California with few first language speakers, a larger number of learners, and active revitalization efforts. Its verbal agreement is sensitive to subject and object, mood, and polarity:

		OBJECT→ SUBJECT↓	1SG	1PL	2SG	2PL	3SG	3PL
NEGATIVE	INDICATIVE	1SG			kín-	ki:k- -ap	ná-	ná-
		1PL			kín-	ki:k- -ap	kín-	kín-
		2SG	ná-	kín- -ap			∅	∅
		2PL	kaná- -ap	kín- -ap			-ap	-ap
		3SG	ná-	kín- -ap	-ap	ki:k- -ap	∅	-ap
		3PL	kaná- -ap	kín- -ap	-ap	ki:k- -ap	-ap	-ap
POSITIVE	INDICATIVE	1SG			nuˀ	ki:kˀ -ap	ni-	ni-
		1PL			nuˀ	ki:kˀ -ap	nuˀ	nuˀ
		2SG	ná-	kínˀ			ʔi-	ʔi-
		2PL	kanáˀ	kínˀ			ku-	ku-
		3SG	ná-	kínˀ	ʔi- -ap	ki:kˀ -ap	ʔu-	ʔu-
		3PL	kanáˀ	kínˀ	ʔi- -ap	ki:kˀ -ap	kunˀ	kínˀ
	OPTATIVE	1SG			nú-	ki:kˀ -ap	kánˀ	kánˀ
		1PL			nú-	ki:kˀ -ap	nú-	nú-
		2SG	ná-	kínˀ			∅	∅
		2PL	kanáˀ	kínˀ			ki:kˀ	ki:kˀ
		3SG	ná-	kínˀ	ʔi- -ap	ki:kˀ -ap	kámˀ	kámˀ
		3PL	kanáˀ	kínˀ	ʔi- -ap	ki:kˀ -ap	kunˀ	kínˀ

Table 1: Karuk verbal agreement (adapted from Bright 1957:64)[3]. Shading indicates object types for which the subject doesn't affect agreement morphology. Abbreviations: 1SG = 1st person singular, etc.

*This paper owes its existence to Line Mikkelsen, who has helped with every step of this project; and to the Karuk elders who identified the set of agreement morphs that this paper is about. Other sources of help include Emily Clem, Amy Rose Deal, Clare Sandy, Peter Jenks, Heidi Harley, and the audiences at UC Berkeley's Syntax and Semantics Circle and WECOL 2017. Thanks.

[1]See also Sappir 2010 for a Distributed Morphology analysis of Karuk's positive indicative agreement paradigm, relying on two case-assigning probes and several Impoverishment rules; Macaulay 1992 and Bank 2010 for analyses involving person-number prominence hierarchies; and White 2015 for a typological perspective.

[2]There's a corpus and dictionary at linguistics.berkeley.edu/~karuk

[3]See Bright 1957:62 for an explanation of the accent notation, and Sandy 2017:159,163-192 for a more nuanced and detailed account of the prosodic properties of Karuk's agreement affixes.

Proceedings of *WECOL 2017*

Karuk has separate negative indicative, positive indicative, and positive optative[4] agreement paradigms, each with an associated set of prefixes encoding information about person and number of the subject and object. There's also one suffix, *-ap*. Notice that for verbs with 1PL or 2PL objects, the subject never affects agreement, whereas for all other object types, the subject can affect agreement – in this sense, we can say that agreement prefers object over subject, and local person and plural over third person and singular. Within Béjar and Rezac's (2009) theoretical framework, this is strong evidence for Karuk verbs having a low cyclic ϕ-probe: such a probe would first Agree with the object, and then, only if the object wasn't maximally ϕ-specified (i.e. wasn't plural number and local person) would the external argument then also Agree. Béjar's (2003) attempt at this kind of analysis of Karuk agreement only accounts for some of the prefixes in the positive indicative paradigm, and relies on incorrectly copied data from Bright's table. Campbell (2012) noticed these issues with Béjar's account, and contributed a much fuller and more correct analysis; but she doesn't cover the negative paradigm's prefixes or the suffix *-ap*, and, as it turns out, also relies on some incorrect data. I spend most of this paper developing a Campbell-style analysis that correctly accounts for the full system. I find that in order to successfully do this, I need to make some tweaks and additions to Campbell's theory of morphosyntax. Within the resulting model – which draws from Deal's (2010, 2015) work on case and agreement – I find that cyclic Agree no longer seems necessary, and I show how an alternative analysis without cyclic Agree could explain some otherwise unexplained facts about Karuk's agreement system.

2 Campbell's analysis clarified, refined, and extended

Like Béjar, Campbell operates under a cyclic Agree framework. Her view of morphology is largely based on Distributed Morphology. After Agree in the syntax, there's Node Insertion in the morphology, whereby the probe terminal gets an AGR sister, with ϕ-features matching the features copied onto the probe under Agree (Campbell 2012:99); next comes Vocabulary Insertion, which selects the Vocabulary item which best (according to the hierarchy in Figure 2.1) expresses the AGR node's feature set.

PERSON	<	PARTICIPANT	<	SPEAKER or ADDRESSEE
∨				
NUMBER	<	NONSINGULAR	<	PLURAL or DUAL
∨				
GENDER	<	ANIMATE	<	FEMININE or MASCULINE

Figure 2.1: According to Campbell (2012:112), expressing ϕ-features follows a crosslinguistic person > number > gender hierarchy of priority, with more specific features preferentially expressed over less specific ones (e.g. PARTICIPANT over PERSON).

Campbell proposes that in Karuk verbs, the head which introduces the external argument is the sole ϕ-probe. She and Béjar call this probe v, but I'll call it Voice, following e.g. Pylkkänen 2008. Campbell ascribes to it the feature specification represented below.

P	<	PART
∨		
N	<	NSG

Figure 2.2: ϕ-specification for Voice in Karuk, following Campbell 2012:145.
abbreviations: Person, PARTicipant, Number, NonSinGular

Notice that this probe can only be fully satisfied by a 1PL or 2PL argument, which is consistent with the shading in Table 1.

[4]Bright's (1957) grammar uses the term *imperative*, but I use *optative*, following Macaulay 1992.

first-cycle (O) morphs			second-cycle (S) morphs			inverse marker
indicative	optative	features	indicative	optative	features	
ná-	ná-	1	ni-	kánˀ	1	-ap
nuˀ	nú-	2	ʔi-	∅-	2	
kínˀ	kínˀ	1PL	nuˀ	nú-	1PL	
ki:kˀ	ki:kˀ	2PL	ku-	ki:kˀ	2PL	
ʔi-	ʔi-	2 / inverse	ʔu-	kám-	3	
			kunˀ	kunˀ	3PL	
			kín-	kín-	3PL / PL	
			ka-	ka-	PL / 1	

Table 2.1: Campbell's (2012:146) agreement Vocabulary for the positive paradigms.

Here's Campbell's Vocabulary table. You can see that the SG feature never appears in the table. Campbell doesn't bother to write down SG because her specificity hierarchy (Figure 2.1) makes the feature unnecessary: there's no danger of e.g. first-cycle *ná-* getting inserted when the object is anything other than 1SG, because when the object is 1PL, the more specific *kínˀ* will be inserted instead. For the same reason, the 3 feature is also unnecessary, so from now on I'll stop writing it. We can also do away with the "P" and "N" in Figure 2.2, giving us a more compact representation of the probe structure:

PART
∨
NSG

Figure 2.3: ϕ-specification for Voice, using simpler notation.

Campbell doesn't attempt to explain why *-ap* occurs in "inverse" contexts; inverse is not a ϕ-feature. The rest of the Vocabulary, which Campbell does provide ϕ-features for, is split in two: there's one Vocabulary for "first-cycle (O)" agreement, and a separate one for "second-cycle (S)" agreement[5].

There's an important implied assumption here: Agree must somehow preserve information about where features came from. That's how first- and second-cycle vocabularies can be available for features from the first and second Agree cycles respectively. When the probe receives first-cycle and second-cycle bundles, it somehow remembers which is which, and somehow directs the morphology to consult the appropriate vocabulary for each.

That's not all. The part of the morphology that consults the second-cycle vocabulary must have access not just to the subject's features, but to features of both the subject and the object, still in neat differentiated bundles[6]. Look at the bottom two rows of Table 2.1: by providing those kinds of feature specifications, Campbell is implicitly assuming that the second-cycle morphology can separately see information about both arguments. And Campbell is right to make that assumption. The difference in prefixes between *kínmah*, 'they see them'; *kunmah*, 'they see her'; and *ʔumah* 'she sees them' is a witness to that.

In the remainder of this section, I improve on Campbell's analysis.

Let's start by thinking about the suffix *-ap*. Here's its distribution in the positive paradigms.

[5]Just to clear up any potential ambiguity as to what "cycle" means: although the main theoretical contribution of Campbell's dissertation is her Cyclic Insert operation, it's clear, based on the annotations "O(bject)" and "S(ubject)", that the relevant cycles here are actually Agree cycles, not Insert cycles.

[6]One way the difference between subject and object might be encoded is if the object's bundle remembers that it has previously Agreed (during the first cycle). This is consistent with a full-feature-exchange model of Agree where the probe's category feature gets copied into bundles it Agrees with (Deal 2010:111, Clem 2017:17). See also Chen's (2018:10) similar proposal that the morphology can differentiate between case features directly received from a case-assigner and case features received secondhand from a case-assigned DP.

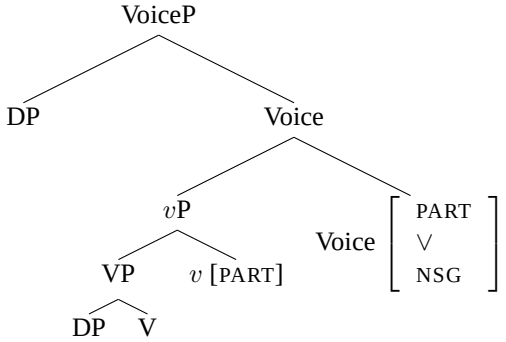
		OBJECT→ SUBJECT↓	1SG	1PL	2SG	2PL	3SG	3PL
POSITIVE	INDICATIVE	1SG			nuˀ	ki:kˀ -ap	ni-	ni-
		1PL			nuˀ	ki:kˀ -ap	nuˀ	nuˀ
		2SG	ná-	kínˀ			ʔi-	ʔi-
		2PL	kanáˀ	kínˀ			ku-	ku-
		3SG	ná-	kínˀ	ʔi- -ap	ki:kˀ -ap	ʔu-	ʔu-
		3PL	kanáˀ	kínˀ	ʔi- -ap	ki:kˀ -ap	kunˀ	kínˀ
	OPTATIVE	1SG			nú-	ki:kˀ -ap	kánˀ	kánˀ
		1PL			nú-	ki:kˀ -ap	nú-	nú-
		2SG	ná-	kínˀ			∅	∅
		2PL	kanáˀ	kínˀ			ki:kˀ	ki:kˀ
		3SG	ná-	kínˀ	ʔi- -ap	ki:kˀ -ap	kám-	kám-
		3PL	kanáˀ	kínˀ	ʔi- -ap	ki:kˀ -ap	kunˀ	kínˀ

Table 2.2: positive agreement morphology. Shaded cells contain *-ap*.

There's reason to think that *-ap*'s insertion is decided by a mechanism somewhat separate from that governing the prefixes' insertion. There's the fact that the suffix is a suffix whereas the prefixes are prefixes, which suggests that they might occupy different syntactic positions. There's also the fact that the distribution of *-ap* is identical between the positive optative and positive indicative paradigms, unlike for the prefixes. And finally, Campbell observes that in the positive paradigms, the contexts *-ap* is inserted in are "inverse" in the sense that the object outranks the subject on a 2 > 1, 3 hierarchy; that's different than the 2, 1 > 3 hierarchy we've seen for prefix insertion.

(It's not solely this hierarchy that determines whether *-ap* gets inserted in positive contexts; otherwise we would falsely predict *-ap* to show up in first person subject, 2SG object forms. But following Campbell, I will assume that the hierarchy plays some role, and will treat the 1 / 2SG contexts as irregular. I'll deal with the regular cases first, and come back to this exception shortly.)

On the one hand, then, *-ap* seems to be an agreement morpheme sensitive to ϕ-features and to "inverseness", which suggests that its insertion is governed by an Agree operation; but on the other hand, it seems to have little to do with the cyclic Agree operation governing prefix insertion. So maybe *-ap* has its own separate Agree probe. Under a 2 > 1, 3 hierarchy, inverse contexts are exactly those in which the object is second person, which means that the part of the grammar operating under this hierarchy only needs to be sensitive to second person, and more specifically, only to second person on the object. We can account for this by adding a probe specified for ADDR in between V and our other probe; the natural thing to call this new probe is "v" (although, just a reminder, for clearness' sake: what Béjar and Campbell call "v" is actually what I've been calling Voice; this new v has nothing to do with that one). Here's what the more articulated structure looks like:



The Vocabulary for v only contains one item: *-ap* ↔ 2.

This correctly predicts the data for the positive paradigms – except for the case I mentioned earlier, where *-ap* doesn't show up when the subject is first person and the object is 2SG. My explanation for this is that only in 1 / 2SG contexts, the v node hosting *-ap* is deleted in the morphology. Arregi and Nevins (2007) propose that in Basque, the presence of 1st and 2nd person features on adjacent nodes is a Person Case Constraint

violation, and that a repair strategy in some dialects is to – to use their terminology – *obliterate* one of the nodes. In Karuk, notice that v only ever Agrees with 2nd person arguments, and Voice only Agrees with subjects when the object isn't local person plural; so the only contexts in which this obliteration rule might be appropriate are 1 / 2SG contexts – i.e. precisely the exceptionally *-ap*-less contexts that this paragraph is about.

Moving on to an unanswered question: how, for the Voice probe, can there can be separate first-cycle 2 and 2 / inverse morphemes (see Table 2.1, reproduced below)? To be clear, the morpheme Campbell says expresses 2 / inverse occurs only when there's *-ap* (in particular, it doesn't show up in 1 / 2SG contexts). But note that Campbell's first-cycle non-inverse 2nd-person morphemes are in fact pronounced the same as Campbell's second-cycle morphemes expressing 1PL. This is true for both positive paradigms – and in fact, if you look back at Table 1, you'll see that the negative paradigm also has a morph with the same distribution (compare optative *nú-*, indicative *nuˀ*, and negative *kín-*).

first-cycle (O) morphs			second-cycle (S) morphs			inverse marker
indicative	optative	features	indicative	optative	features	
ná-	ná-	1	ni-	kánˀ	1	-ap
nuˀ	nú-	2	ʔi-	∅-	2	
kínˀ	kínˀ	1PL	nuˀ	nú-	1PL	
ki:kˀ	ki:kˀ	2PL	ku-	ki:kˀ	2PL	
ʔi-	ʔi-	2 / inverse	ʔu-	kám-	3	
			kunˀ	kunˀ	3PL	
			kín-	kín-	3PL / PL	
			ka-	ka-	PL / 1	

Table 2.1 again: Campbell's (2012:146) agreement Vocabulary for the positive paradigms.

This new observation makes it seem like perhaps Campbell's first-cycle 2nd-person morpheme actually doesn't exist as a separate morpheme, and that instead it's the second-cycle 1PL morpheme that's inserted in those contexts[7]. There are two subject/object configurations in which Campbell's first-cycle 2nd-person morpheme gets inserted: 1PL / 2SG and 1SG / 2SG. One of those configurations has a 1PL subject, so that seems to fit with our proposed reanalysis; and as for the other configuration, we already know that the second-cycle morphology can see features of both the subject and the object, so it's plausible that 1SG+2SG combine to create an inclusive we, which gets expressed via the second-cycle 1PL morpheme. At this point, another question arises, though: why is this morpheme not inserted in 1SG / 3SG and 1SG / 3PL contexts? The first case has a simple explanation: a 3SG DP has no ϕ-features specific enough to be visible to the probe, so unlike for 1SG / 2SG, where we end up with 1+2=1PL, for 1SG / 3SG there's just a lonely 1=1SG. As for the case where the object is 3PL, the probe will Agree with it; but I claim that 1+PL≠1PL. As evidence for this, I turn to some of Karuk's DP morphology. Below, you can see that the plural marker *kun* is unable to combine with 1st-person markers[8]. I take this as evidence that in Karuk, 1PL cannot be decomposed into 1+PL[9].

	SG	PL
1	nani-/nini-	nanu-/nunu-
2	mi-	mikun-
3	mu-	mukun-

Table 2.3: Personal possessive prefixes (Bright 1957:56).

	SG	PL
1	náa	núu
2	íim	iimkun
3	úum	uumkun

Table 2.4: Personal pronouns (from the dictionary at http://linguistics.berkeley.edu/∼karuk/).

[7]This has been proposed before by Gluckman (2016:26)

[8]Line Mikkelsen alerted me to this.

[9]But see section 3 for a perhaps more plausible explanation of the 1PL prefixes' distribution.

Now there's only one 1st-cycle 2nd-person morpheme remaining: the "2 / inverse" morpheme. What if we were to just recast this morpheme as simply a 2nd-person morpheme, without having to say anything about inverse? That would correctly predict its insertion in the places it's attested, of course, but we also would predict that it would be inserted in 1 / 2SG contexts, which it isn't. Under our current set of expectations, we would expect two prefixes here: the first-cycle 2 (= "2 / inverse") prefix, and the second-cycle 1PL prefix. The fact that instead only one prefix shows up seems to reveal a rule in the morphology that says there's only room for one agreement prefix. There's independent evidence for this rule: 'they see you.SG' is pronounced *ʔimahap*; the second-cycle PL prefix *kunˀ* fails to be inserted, even though it would expone features not exponed by *ʔi-* ↔ 2. A similar thing seems to happen in 1 / 2SG contexts: the 1PL prefix outcompetes the 2 prefix for a single slot. The assumption that 1PL takes priority over 2 is consistent with the rest of the data: under my analysis, 1st person is the only feature for which whenever Agree encounters it, it gets expressed in the agreement morphology (the one case where there's an unexpressed 1st person feature is the 1 / 2PL case, but in that case Voice stops probing after one cycle, so there's no opportunity for the 1 feature to even be seen by Agree); and we also saw earlier that in 1 / 2SG contexts, of the two candidate nodes to be obliterated, the morphology chooses to delete v, i.e. the non-1st-person node.

The problem with proposing a one-prefix-only rule is that Campbell analyzes Karuk agreement as having one case of double prefixation: she claims that *kaná*, the prefix that occurs in plural-subject, first-person-singular object constructions, is decomposable into the first-cycle first-person morpheme *ná* and the second-cycle morpheme *ka*. But you might remember that at the beginning I mentioned that Campbell's analysis relied on incorrect data – this is where that comes in. In the positive paradigms, *kanáˀ* is written with an accent mark on the hyphen; this denotes a phonological effect called post-accenting[10], and the first-cycle prefix *ná-* isn't post-accented, so Campbell's decompositional analysis is incorrect.[11] In any case, the decompositional analysis was unmotivated, since *ka* never occurs in any other context; we can analyze *kanáˀ* as a single second-cycle PL / 1 morpheme. We also have to explain why *ná-* is not inserted in plural-subject constructions; we can account for that subject-dependency by recasting it as a second-cycle morpheme exponing / 1.

Before we move on to the negative paradigm, here's the positive prefixes in a table:

first-cycle (O) morphs			second-cycle (S) morphs		
indicative	optative	features	indicative	optative	features
			ni-	kánˀ	1
ʔi-	ʔi-	2	ʔi-	∅-	2
kínˀ	kínˀ	1PL	nuˀ	nú-	1PL
ki:kˀ	ki:kˀ	2PL	ku-	ki:kˀ	2PL
			ʔu-	kám-	
			kunˀ	kunˀ	PL
			kín-	kín-	PL / PL
			kanáˀ	kanáˀ	PL / 1
			ná-	ná-	/ 1

Table 2.5: Positive Vocabulary for Voice.

In the negative paradigm, *-ap* occurs everywhere positive *-ap* does, and also in some other environments; in those other environments, there's always a plural argument. I claim, therefore, that there are actually two different suffixes that happen to both be pronounced *-ap*[12]: the universal, object-dependent *-ap*$_2$ comes from a probe on v, and expones 2nd person, as we've seen; and the negative-only *-ap*$_{neg}$ is some sort of plural marker which only appears in negative contexts. We see in Table 2.6 that in the negative paradigm, *-ap* surfaces:

- everywhere it would appear in the positive paradigms
- when the subject is 2nd person and one of the arguments is a plural participant
- when the subject is 3rd person and one of the arguments is plural.

[10] See Sandy 2017:159.

[11] Thanks to Clare Sandy for helping me identify this as an issue.

[12] This has been suggested before by Macaulay (1992:194).

		OBJECT→ SUBJECT↓	1SG	1PL	2SG	2PL	3SG	3PL
NEGATIVE	INDICATIVE	1SG			kín-	ki:k- -ap	ná-	ná-
		1PL			kín-	ki:k- -ap	kín-	kín-
		2SG	ná-	kín- -ap			∅	∅
		2PL	kaná- -ap	kín- -ap			-ap	-ap
		3SG	ná-	kín- -ap	-ap	ki:k- -ap	∅	-ap
		3PL	kaná- -ap	kín- -ap	-ap	ki:k- -ap	-ap	-ap
POSITIVE	INDICATIVE	1SG			nuˀ	ki:kˀ -ap	ni-	ni-
		1PL			nuˀ	ki:kˀ -ap	nuˀ	nuˀ
		2SG	ná-	kínˀ			ʔi-	ʔi-
		2PL	kanáˀ	kínˀ			ku-	ku-
		3SG	ná-	kínˀ	ʔi- -ap	ki:kˀ -ap	ʔu-	ʔu-
		3PL	kanáˀ	kínˀ	ʔi- -ap	ki:kˀ -ap	kunˀ	kínˀ
	OPTATIVE	1SG			nú-	ki:kˀ -ap	kánˀ	kánˀ
		1PL			nú-	ki:kˀ -ap	nú-	nú-
		2SG	ná-	kínˀ			∅	∅
		2PL	kanáˀ	kínˀ			ki:kˀ	ki:kˀ
		3SG	ná-	kínˀ	ʔi- -ap	ki:kˀ -ap	kám-	kám-
		3PL	kanáˀ	kínˀ	ʔi- -ap	ki:kˀ -ap	kunˀ	kínˀ

Table 2.6: Karuk agreement, with cells containing *-ap* shaded.

The first bullet point is covered by the universal, 2nd-person-exponing *-ap*$_2$. I claim that the other *-ap*, *-ap*$_{neg}$, is responsible for the other two bullet points. I propose that *-ap*$_{neg}$ is inserted at Neg, which probes downward, first Agreeing with the subject and then possibly with the object too. I follow Deal's (2015) proposal that probes' *interaction criteria* – the criteria that decide whether or not the probe will Agree with a node it encounters during its search – and *satisfaction criteria* – the criteria that decide whether after Agreeing with a node, the probe will stop searching or keep searching – are not necessarily the same. But I make a substantial addition to Deal's model. To correctly capture the distribution of *-ap*$_{neg}$, I assume that just as full satisfaction of a probe triggers full shutdown of interaction, partial satisfaction of a probe can trigger partial restriction of the probe's interaction criteria. In particular, I assume that Karuk's Neg probe has satisfaction criteria {PARTICIPANT, SPEAKER} and is initially willing to interact with any DP, but after it Agrees with a second person subject, the only remaining feature it needs for satisfaction is SPEAKER, and so from that point on it'll only interact with nodes containing a SPEAKER feature.

subject	probe behavior
1st person	always Agrees with the subject; never Agrees with the object
2nd person	always Agrees with the subject; Agrees with the object only if it's 1st person
3rd person	always Agrees with the subject and the object

Table 2.7: Neg probe behavior by subject type.

The Vocabulary for this probe consists of just two items: first-cycle ∅ ↔ 1 and second-cycle *-ap* ↔ PL. This correctly predicts the distribution of *-ap*.

One last thing about *-ap*: my analysis predicts double insertion of *-ap* in 3PL-subject, second-person-object contexts. But the fact that *-ap* is only pronounced once in those contexts is not surprising from a language internal perspective: elsewhere in Karuk, haplology prevents pronouncing more than one of successive identical morphs (Bright 1957:122,125,126).

For the negative-paradigm prefixes, the facts are consistent with the following Vocabulary for the Voice probe:

1st cycle	2nd cycle
kín- ↔ 1PL	ná- ↔ 1
ki:k ↔ 2PL	kín- ↔ 1PL
	kaná- ↔ PL / 1

Table 2.8: Negative Vocabulary for Voice.

In the second-cycle column, only feature sets containing the 1st person feature get a pronounced exponent. We just saw that for the Neg probe, it's the opposite: the 1st person feature has an unpronounced exponent on that probe. So Voice's second-cycle Vocabulary and Neg's *-ap* nicely complement each other – but it's not a strict complementarity, since Neg's *-ap* co-occurs with the second-cycle morpheme *kaná-* in plural-subject, 1SG-object contexts. The non-strictness of the complementarity is evidence in favor of my analysis where Neg's *-ap* and second-cycle prefixes depend on separate syntactic processes, and their surface complementarity is the expression of a Vocabulary that has perhaps over time undergone reduction in response to pressure to be efficient and limit redundancy. The extra probe in negative contexts lightens the phonetic workload of the Voice probe.

That covers Karuk's entire agreement system. But along the way, we developed a model of grammar where:

- there's multiple probes
- a single AGR node can contain differentiated feature bundles from multiple nodes
- probes are relativized but Agree consists of full feature exchange
- probes' interaction behavior depends on their degree of satisfaction
- the morphology can mediate between multiple candidate morphemes at one syntactic node
- the morpholgy can obliterate syntactic nodes.

At this point, the theory is powerful enough that cyclic Agree seems superfluous.[13]

3 Towards an analysis without cyclic Agree[13]

To wrap up, let's think about the implications of the work in section 2, both for our understanding of agreement in general and for our understanding of Karuk agreement specifically.

Cyclic Agree is a mechanism where probes from two different heights in the syntactic structure – one high enough to see the subject, the other lower and only able to see the object – can have their features wind up together on one syntactic node. In a model where the external argument is introduced by the verbalizing head v, cyclic Agree is perhaps the simplest such mechanism. But in a model with a more complex verb phrase structure, where v is distinct from Voice, we can easily construct such a mechanism without recourse to cyclic Agree, by placing separate non-cyclic probes on v and Voice (with head movement and/or Agree causing v's features to move up to Voice). And we've seen in section 2 that the more complex verb phrase structure and the separate probes on v and Voice are useful stipulations even if Voice is a cyclic probe. Which means that for cyclic Agree to be a useful stipulation, it needs to do something more than what I described in the first sentence of this paragraph.

Cyclic Agree does do something more: it derives direct/inverse alternations along specificity hierarchies, where the object is preferred over the subject for control of agreement provided it's ϕ-specific enough. But with separate v and Voice probes, we can derive the same kind of object-preference, using the type of full feature exchange model adopted in section 2 (page 3): if the object's bundle Agrees with both v and Voice, then the morphology at Voice will prefer, given equal ϕ-specificity of subject and object, to expone the object's bundle, because the object's bundle will be more specific in that it'll carry a v feature. And in fact, the model of Agree developed in section 2 provides another, completely different, purely syntactic way to derive direct/inverse hierarchy effects: under my "partial satisfaction" model, a high probe will start by Agreeing with the subject, and then will only Agree with the object if the object is sufficiently ϕ-specific compared to the subject.

This model of Agree can explain the distribution of *-ap* in the negative paradigm, as we've seen in section 2, which can't be explained by cyclic Agree. It can also explain why 1PL prefixes fail to be inserted in 1SG / 3

[13] By "cyclic Agree", I just mean the interleaving of Merge and Agree. I'm not claiming that we can get rid of other kinds of cyclicness – for instance, I still want to assume that interaction criteria are updated after each interaction, and that there's separate first- and second-cycle Vocabularies, as in my proposed Neg probe.

contexts: if those prefixes come from a high probe with satisfaction criteria {PARTICIPANT, ADDRESSEE, SPEAKER}, for instance, then when there's a 1st person subject, 2nd person objects will be Agreed with but 3rd person objects won't.[14] A benefit of this analysis is that it correctly predicts a fact viewed as coincidental under the analysis presented in section 2: Karuk's agreement morphology isn't sensitive to 3rd person objects' features except when the subject is also 3rd person.

		OBJECT→ SUBJECT↓	1SG	1PL	2SG	2PL	3SG	3PL
NEGATIVE	INDICATIVE	1SG			kín-	ki:k- -ap	ná-	ná-
		1PL			kín-	ki:k- -ap	kín-	kín-
		2SG	ná-	kín- -ap			∅	∅
		2PL	kaná- -ap	kín- -ap			-ap	-ap
		3SG	ná-	kín- -ap	-ap	ki:k- -ap	∅	-ap
		3PL	kaná- -ap	kín- -ap	-ap	ki:k- -ap	-ap	-ap
POSITIVE	INDICATIVE	1SG			nuˀ	ki:kˀ -ap	ni-	ni-
		1PL			nuˀ	ki:kˀ -ap	nuˀ	nuˀ
		2SG	ná-	kínˀ			ʔi-	ʔi-
		2PL	kanáˀ	kínˀ			ku-	ku-
		3SG	ná-	kínˀ	ʔi- -ap	ki:kˀ -ap	ʔu-	ʔu-
		3PL	kanáˀ	kínˀ	ʔi- -ap	ki:kˀ -ap	kunˀ	kínˀ
	OPTATIVE	1SG			nú-	ki:kˀ -ap	kánˀ	kánˀ
		1PL			nú-	ki:kˀ -ap	nú-	nú-
		2SG	ná-	kínˀ			∅	∅
		2PL	kanáˀ	kínˀ			ki:kˀ	ki:kˀ
		3SG	ná-	kínˀ	ʔi- -ap	ki:kˀ -ap	kám-	kám-
		3PL	kanáˀ	kínˀ	ʔi- -ap	ki:kˀ -ap	kunˀ	kínˀ

Table 3: Karuk agreement.

The idea that Karuk's agreement prefixes might originate from multiple probes instead of one cyclic probe is intriguing for a few reasons. The syncretism[15] of prefixes across paradigms when the object is 1SG, 1PL, or 2PL suggests that those prefixes might be inserted syntactically lower than the other, mood- and polarity-dependent, prefixes. On the other hand, Sandy (2017:163-192) has recently argued that some of Karuk's agreement prefixes are within a prosodic word boundary and others aren't, and that these patterns can't be fully explained under the cyclic Agree analysis; an analysis where those two classes of prefixes come from separate probes seems worth attempting. There's also the fact that the prefixes which negative *-ap* cooccurs with are all shared across all three paradigms; maybe this is evidence that the prefixes unique to the negative paradigm are able to block *-ap*, which might mean that the uniquely negative prefixes come from the same probe as negative *-ap*, whereas the shared prefixes come from a lower probe.

References

Arregi, Karlos and Andrew Nevins. 2007. Obliteration vs. Impoverishment in the Basque g-/z- Constraint. *Penn Working Papers in Linguistics* 13.1:1-14.

Bank, Sebastian. 2010. Complex Scales in Multiargument Agreement. *2 in Agreement*, ed. by Sebastian Bank, Doreen Georgi, and Jochen Trommer, 65-92. Insitut für Linguistik, Universität Leipzig.

Béjar, Susana. 2003. Phi-Syntax: A Theory of Agreement. Dissertation, University of Toronto.

Béjar, Susana and Milan Rezac. 2009. Cyclic Agree. *Linguistic Inquiry* 40:35-73.

Bright, William. 1957. The Karok Language. Berkeley: University of California Press.

Campbell, Amy Melissa. 2012. The Morphosyntax of Discontinuous Exponence. Dissertation, UC Berkeley.

Chen, Yitong. 2018. Case Stacking as a Post-syntactic Operation. These proceedings.

[14] Compare this analysis to the one proposed on page 5.

[15] Syncretism up to post-accenting in the negative paradigm.

Clem, Emily. 2017. Amahuaca ergative as agreement with multiple heads. Manuscript, UC Berkeley.

Deal, Amy Rose. 2010. Ergative case and the transitive subject: a view from Nez Perce. *Natural Language and Linguistic Theory* 28:73-120.

Deal, Amy Rose. 2015. Interaction and satisfaction in ϕ-agreement. *Proceedings of NELS 45*, ed. by Thuy Bui and Deniz Özyıldız, 1:179-192.

Gluckman, John. 2016. Decomposing Number Features. Manuscript, UCLA.

Macaulay, Monica. 1992. Inverse Marking in Karuk: The Function of the Suffix -ap. *International Journal of American Linguistics* 58:182-201.

Pylkkänen, Liina. 2008. Introducing Arguments. Cambridge, MA: MIT Press.

Sandy, Clare Scoville. 2017. Prosodic Prominence in Karuk. Dissertation, UC Berkeley.

Sappir, Michael. 2010. Hierarchies and Portmanteaux of Karuk Pronominal Affixes with Distributed Morphology and without Hierarchies or Portmanteaux. *2 in Agreement*, ed. by Sebastian Bank., Doreen Georgi, and Jochen Trommer, 1-22. Insitut für Linguistik, Universität Leipzig.

White, Whitney. 2015. Typological Canonicality of Karuk Agreement. Honors thesis, UC Berkeley.

The Grammar of Space in Northern Paiute Narrative Domains

Hannah Masson and Christina Trouten
Boise State University

1 Introduction

Northern Paiute postpositions and demonstratives play important roles in expressing spatial, directional, temporal, instrumental, comitative, and simulative semantic domains. In this paper we explore the grammatical functions of postpositions and demonstratives in Northern Paiute (Numic: Uto-Aztecan) to show how they interact with each other and with other parts of the grammar in order to express movement through space and time, event location, and to track reference of peripheral arguments in narrative. We will begin with a brief discussion of postpositions in the language, followed by a preliminary analysis of a subset of demonstratives that serve to track location in discourse, and ending with some hypotheses regarding the interaction of the two systems in the grammar of space.

2 Postpositions in Northern Paiute

Postpositions are a type of adposition. An adposition can be defined as "a word class that occurs with a noun phrase and that indicates the grammatical, spatial, temporal, or logical relationship of the noun phrase to another element of the clause" (Genetti 2014:596). As the name implies, postpositions come after their complements. They can be compared to prepositions in English as both express similar kinds of information---spatial, directional, temporal, instrumental, comitative, or simulative relationships, differing only with regard to their placement.

In Northern Paiute postpositions show the function of a participant in a clause "as location, instrument, or some other peripheral argument" (Thornes 2003:170). They most often appear as suffixes to nouns, demonstratives, and pronominal bases. The system is elaborate, consisting of nearly 30 forms. These are in multiple strata of free and bound forms that are frequently combined and occur in particular orders. Some may only come after another postposition; some only manifest with demonstratives or pronominal bases. Many forms may express multiple meanings depending on the context and to what other postpositions they attach. Their various distributional patterns reflect historical influences of grammaticalization and lexicalization of stative verbs and relator nouns (Thornes 2003:195).

There are two basic nominal construction types that involve postpositions in Northern Paiute. The most common is the simple "noun plus postposition" construction expressing basic spatial relationships, which involves simple suffixation or juxtaposition of a nominal with a postposition or postpositional complex (Thornes 2003:222). An example of the noun plus postposition construction expressing a spatial relationship is as follows:

(1)	*umɨ*	*yaisi*	*mii*	*una–tami*	*mia–si*	
	3pl	then	QUOT	over.there–TOWARD	go.SG–SEQ	
		waa–tuha–'yu		*tammi*	*tika–ga*	*mii*
		juniper–UNDER–NOM		1pl.INCL	eat–TRNSL	QUOT

'They then said, "We're going over there under the juniper to eat." (NK:Root-Digging Time, ln. 57)[1]

[1] Examples from narrative texts (Thornes, nd.) are indicated with speaker initials, text title, and line number. All examples in this study came from a single, lengthy autobiographical narrative, as told by the late Nepa Kennedy (b. 1918, d. 2010) and recorded by Tim Thornes on April 1, 1999 on the Burns Paiute Reservation, Oregon.

Proceedings of *WECOL 2017*

Here there are two postpositions: one (-tami, 'toward') attaches to a demonstrative (una, 'over there'). The other (-tuha, 'under') attaches to a noun root (waa-, 'juniper'). The first postposition helps the demonstrative meaning 'over there' to be more specific by adding directional information relating to the juniper. Using 'una' normally would reference a general area away from the speaker. Adding the postposition '-tami' creates a more exact reference point, directing the interlocutor's focus toward another object mentioned in the same sentence. The second postposition, '-tuha' expresses another spatial relationship with the juniper, but in this case it is not a direction but a place that the speaker and his interlocutor will occupy while they eat.

The next construction type involving postpositions is the pronoun copy construction, where a postposition is bound to a pronominal or demonstrative base that co refers to an independent nominal that appears elsewhere in the clause. The first postposition in the above example is an example of the pronoun copy construction as it attaches to a demonstrative and co refers to the juniper later mentioned in the sentence. Another example of this type of construction is as follows:

(2)	*oo =tia'a*	*pi-kwai-tu*	*ni= kima-kwai*	*nimmi,*	*mi= nobi*
	DEM =thusly	RESTR-AREA-TO	our= come-AREA	we.EXCL	our= house

'Then also we went back to our digging place, too, and' (NK:Root-Digging Time, ln. 77)

In this example, the third person restrictive pronoun base (pi-) is used with the postpositions -(k)wai and -tu ('at' and 'to', respectively) to reference the head of a relative clause (the noun 'nobi', meaning house/home) as a kind of "it" place.

Another function of postpositions in Northern Paiute is called locative nominalization. This is the ability to create a noun location by using a postposition with a verb. In this case the postposition attaches to a verb to create a nominal place where the verb is carried out. An example is as follows:

(3)	*yaisi*	*tiwau*	*nimmi*	*pinau–su*	*mi= tihona–wai–tu*
	then	also	1pl.EXCL	back–ADV	1pl= dig.roots–AREA–TO

	mia–si	*tiwau*
	go.SG–SEQ	also

'Then also we went back to our digging place, too, and' (NK:Root-Digging Time, ln. 56)

In this example the verb is 'tihona' (meaning 'dig roots') and two postpositions attach to it (-(k)wai, -tu, meaning 'at' and 'to', respectively). This references the place where digging roots occurs.

3 Demonstratives in Northern Paiute

The other important component in the Northern Paiute grammatical construction of space is the demonstrative class. While there are many demonstratives in the Northern Paiute language, we will focus our attention on five that are used mainly in the expression of distal-spatial relationships. These are 'owi', 'obi', 'oo', 'oi', and 'una'. These spatial demonstratives serve three evident purposes. Firstly, a demonstrative may reference a non-specific place that is also not literally locative as a discourse marker. This is most often accomplished with the demonstrative 'oo' which (very roughly) translates to "out there". Next, they may refer to a literal location that has either been mentioned previously in the discourse, or is just about to be mentioned.

The final function these spatial demonstratives serve in discourse is to mark general, imprecise locations. This is most commonly seen when the speaker is referring to a location where an event has occurred or action has been performed, but has not named a specific place or geographic location. This is illustrated in the following example:

(4) *o'o–ti–u* *umɨ* *tɨhona–kwɨ* *mii* *ɨni–na ka*
DIST–LOC–FOC 3pl dig.rootsFUT QUOT say–PTCP PTCL
'Right there is where they would dig, (they) were saying.' (NK:Root-Digging Time, ln. 21)

In this case, the demonstrative 'oo' is used to reference the geographic location where the root digging would take place. The place referenced is not near the speaker, and has not been specifically named in the discourse already as a specific location (such as a hill or a field) making the use of 'oo' as a location marker acceptable. In cases where the location being referenced has already been specifically mentioned, 'owi' or 'obi' (covered later on) may be more appropriate.

There is some functional overlap among the five demonstratives examined for the purposes of this paper. For example, both 'oo' and 'una' can be used in the context of example one. What determines which form appears remains puzzling. We initially assumed that the two demonstratives were interchangeable. However, we have yet to find evidence suggesting that 'una' can be used in the third above-mentioned context, that is, to reference a literally locative but non-specific place. Of all five demonstratives examined, 'oo' was the only one to appear in this context in the data set. It appears to be the most versatile of the five as well, as it is also used to indicate place of action, and in cases where the place being referenced has not been explicitly named already in the discourse but is about to be, as in the following:

(5) *o'o* *ka= yappa–wɨnɨ–'ɨ* *mii* *tɨ= ni'a–na*
DIST ACC= ipos–stand.SG–NMR QUOT LOGO= call–PTCP
'Out there "Ipos–Standing" is its name;' ' (NK:Root-Digging Time, ln. 18)

Here, 'oo' is used as a reference to "Ipos-standing", a place-name that has yet to be explicitly mentioned in the discourse. In the majority of cases in which demonstratives show up in this context, the place name is given immediately after the use of the demonstrative 'oo', again suggesting that it serves an emphatic function as a discourse marker.

Demonstratives 'owi' and 'obi' also share a common environment which makes it unclear how they are distinguished from one another. Both can be used to refer anaphorically to another demonstrative. There may be a hierarchical order in which demonstratives appear in these situations, as seen below.

(6) *obi–u* *owi–u* *nɨmmi* *o'o–ti–u =tia'a*
in.there–FOC there–FOC 1pl.EXCL DIST–LOC–FOC =so/thus
'that's where we (went); just there.' (NK:Root-Digging Time, ln. 19)

The above example makes thorough use of the demonstrative system and allows us to see the proposed hierarchical structure at work. The previous line in the narrative (example 5) begins with the demonstrative that sees the most general use, 'oo'. This is referenced anaphorically by obi and immediately referred to again with 'owi'. Compared against the English gloss, though a one-to-one translation is never truly possible, it could be said that 'obi' is the 'that' in the statement and 'owi' is the 'where'. At the end of the statement, 'oo' is used again for emphasis, completely the discourse marker function mentioned earlier in this paper. The focusing suffix '–u' appears on each of the three demonstrative forms in this line of the narrative.

The demonstrative 'oi' is the only one in the set that doesn't show significant functional overlap with the others. It seems to be used most often in situations where the speaker is referring to the goal of their movement, as opposed to simply referencing a static position.

(7) *yaisi nɨmmi yaisi oi–tu tɨ– mahi–ga*
then 1pl.EXCL then there–TO APS–investigate–TRNSL
'Then we then went over there to investigate;' (NK:Root-Digging Time, ln. 19)

Here we see 'oi' used as a demonstrative pronoun for a place moved towards within the narrative, rather than simply a place being referenced at a discourse level. Unlike the other demonstrative pronouns in our set, 'oi' often appears in conjunction with postpositions, such as 'tu' seen above as a post position meaning 'towards'. 'Oi' also appears with the locative postposition '-wai', seen above in section 2.

4 Conclusion

Further research is needed to explore broad distributional patterns of the demonstratives as well as why specific forms of postpositions and demonstratives are chosen over other forms that express similar information, as the differences in expressed meaning are often extremely subtle, and overlap in functionality is common. Postpositional distribution patterns are also worth further investigation as there are multiple strata and the forms used are interdependent. A deeper analysis of the postpositional forms that are used exclusively with demonstratives would be insightful as well as this would highlight more fully the ways in which these parts of the grammar come together to specify spatial information.

5 References

Genetti, Carol. 2014. How Languages Work: An Introduction to Language and Linguistics. Cambridge: Cambridge University Press.
Thornes, Tim. 2003. A Northern Paiute Grammar with Texts. Ph.D. dissertation. University of Oregon: Eugene.
Thornes, Tim (ed.) no date. Wadateka'a Nadeguyunga: Harney Valley Stories. Unpublished manuscript.

Measure Phrases in Spanish Second Language Learners and Heritage Speakers

Mythili Menon and Rachel Showstack
Wichita State University

1 Introduction

The linguistic repertoires of second language (L2) learners and heritage speakers (HS) have been compared, in the recent past, to better understand the processes of language acquisition, the possibility of transfer from the dominant language, the differing communicative practices, and mental grammars of these two broadly defined groups (e.g. Montrul, 2010; Showstack, 2017). While research has demonstrated that the Spanish of adult HS differs from the Spanish of monolinguals of the same age group, and that there is a great deal of diversity within the HS group, these differences need further exploration. In addition to contributing to formal linguistics, such studies are also important to heritage language (HL) pedagogy because they show how the pedagogical needs of HS differ from those of L2 learners (Bayram, Prada, Pascual y Cabo, and Rothman, 2016). In a study of Spanish clitics, Montrul (2010) found that HS had clear advantages over L2 learners in some areas but concluded that both groups showed effects of English in some areas, particularly at the interface between syntax and semantics/pragmatics. While it is difficult to determine whether a given linguistic feature does indeed exhibit crosslinguistic influence (Silva-Corvalán, 1993) and scholars in applied linguistics caution against classifying similarities to English in bilingual speech as a deficit (García, 2009; Showstack, 2017), further research is needed on the mental grammars and linguistic production of L2 learners and HS at the syntax/semantics interface to understand the effects of different bilingual learning contexts on the linguistic repertoires of speakers from these two groups.

For Spanish HS and L2 learners, one syntactic/semantic feature that merits further investigation is measure phrases (e.g. five inches, two pounds), because syntax/semantics research suggests that they are realized differently in Spanish and English. Measure phrases are often taken to be modifiers of gradable adjectives (such as *tall*). In degree expressions, they measure an amount or a difference on a scale. In English, measure phrases appear with both the positive form of the adjective or the comparative form of the adjective as in (1)[1].

(1) a. Mary is six feet tall.
b. Mary is six inches taller.

In (1a), the adjective *tall* is in the positive form. The measure phrase *six feet* measures the height of Mary. In (1b), the adjective *taller* is in the comparative form, with the *–er* marker attaching onto the positive form of the adjective. Here, the measure phrase measures the difference in height between Mary and the target of comparison (which is implicit in this case). We will refer to these constructions as the non-comparative form and the comparative form respectively.

Languages differ on the basis of whether they allow measure phrases with both the comparative and non-comparative forms. Contrary to English, research in syntax/semantics posits that Spanish measure phrases can only appear with the comparative form of the adjective (Sáez del Álamo, 1997; Bosque, 1998;

* The authors thank Alfredo García Pardo for help with the experimental stimuli and the audience at WECOL 2017 for helpful comments.

[1] This distinction between measurements is often referred to as the absolute measurement versus differential measurement. Languages vary internally and cross-linguistically between these two kinds of measurements (see, Schwarzschild, 2005).

Proceedings of *WECOL 2017*

Schwarzschild, 2005; and, Eguren and Pastor, 2014), as shown below.

(2) a. Juan es un metro más alto que Pedro (Eguren and Pastor, 2014 (1a, b))
John is a meter more tall than Peter
'John is one meter taller than Peter.'

b. *Juan es un metro alto
John is a meter tall
Intended: John is one meter tall

In (2), the measure phrase *un metro* is compatible only with the comparative form of the adjective *más alto*. Without the comparative marker *más*, the combination is ungrammatical, as shown in (2b). Thus, measure phrases are incompatible with non-comparative adjectives in Spanish. However, it has been noted in the literature that some types of adjectives do allow measure phrases with non-comparative adjectives in Spanish. These adjectives have a lower closed scale, according to the typology of Kennedy and McNally (2005)[2].

(3)	A typology of scale structure		Example
	a. (Totally) open scale	○----------○	*tall*
	b. Lower closed scale	●----------○	*dirty*
	c. Upper closed scale	○----------●	*clean*
	d. (Totally) closed scale	●----------●	*full*

Lower closed scale adjectives such as *dirty* are compatible with the degree modifier 'slightly' and incompatible with the degree modifier 'perfectly'. This is due to the fact that 'slightly' picks out the minimum degree on the adjective's scale whereas 'perfectly' picks out the maximum degree. In lower closed scale adjectives, the maximum degree is open and therefore the incompatibility.

Languages that do not generally allow measure phrases combining with non-comparative adjectives (Spanish, Russian, Korean), allow the acceptability of measure phrases with lower closed scale adjectives (Sawada and Grano, 2011), as shown in the following example from Spanish.

(4) Esta varilla está doblada noventa grados (Sawada and Grano, 2011: (8))
This rod is bent ninety degrees
'This rod is ninety degrees bent.'

As seen in (4), the measure phrase 'noventa grados' modifies the lower closed scale non-comparative adjective 'bent'. Note that the gradable predicate in this sentence is deverbal, however the semantics of the predicate does not change irrespective of the morphology and it is the semantics that is crucial for the measure phrase. In addition, Eguren and Pastor (2014) has noted that some measure phrases can combine with bare adjectives in the [MP + *de* + A] pattern, where *de* is a prepositional copula heading a small clause, as shown in (5).

(5) Una valla de piedra de [dos metros de alta] (Eguren and Pastor, 2014: (5))
A $\text{fence}_{\text{FEM}}$ of stone of two meters of high-FEM
'a two meters high stone fence'

In the above example, the measure phrase 'dos metros' combines with the prepositional copula *de* and the adjective 'alta', again contrary to the assumption that measure phrases cannot combine with a non-comparative adjective. To summarize this section, research suggests that English and Spanish allow measure phrases with comparative adjectives, thereby allowing the differential measurement reading, but

[2] These adjectives are also referred to as absolute adjectives versus relative adjectives (see, Kenney and McNally 2005, Toledo and Sassoon 2011, a.o).

Spanish does not allow measure phrases with non-comparative adjectives, thereby disallowing the absolute measurement reading. However, in Spanish, certain non-comparative adjectives combine with measure phrases. These adjectives belong to the lower closed scale adjectives. In addition, measure phrases can also combine with the prepositional copula *de* and together modify a non-comparative adjective. This is summarized below in table 1 below.

	English	**Spanish**
Allows measure phrases with comparative adjectives	√	√
Allows measure phrases with all non-comparative adjectives	√	x
Allows measure phrases with some non-comparative adjectives (lower closed scale, with prepositional *de*)	na	√

Table 1: Summary of existing syntax/semantics literature on English and Spanish Measure Phrases

The literature highlighting these judgments comes from native speaker intuitions (Schwarzschild, 2005; Sawada and Grano, 2011) and corpus data (Eguren and Pastor, 2014). No study has looked at the acceptability of measure phrases with adjectives with different Spanish populations, given that the existing theoretical literature makes contradicting predictions on the acceptability of measure phrases.

In this paper, we present results from an acceptability judgment study and a usability study we conducted, looking at the acceptability and usability of measure phrases in Spanish with populations whose use of measure phrases has not been well studied, namely second language learners and heritage speakers. We chose to study these populations to find out how much they show similarities to English at the syntax/semantics interface. Our results, in the acceptability rating study, show that L2 learners of Spanish prefer measure phrases with comparative adjectives and prefer measure phrases without the prepositional copula *de*. In contrast, our heritage speakers of Spanish had no preference between comparative and non-comparative adjectives. However, they preferred adjectives with the prepositional copula *de* to those without it. Our results, in the usability study, show that both L2 learners and HS rated adjectives without *de* as the sentences they use the least in everyday speech. Both groups also preferred to use adjectives with *de* the most.

The rest of the paper is structured as follows. In the next section, we provide a quick overview of the syntax and semantics of adjectives and measure phrases. In section 3, we introduce our study and the methodology, section 4 presents the results of the study, section 5 is the general discussion and section 6 lays out the plan for future studies.

2 Adjectives and Measure Phrases

In this section, we briefly provide an overview of the previous analyses of adjectives and measure phrases in the theoretical literature. There are many competing analyses for the semantics of adjectives. Under the standard analysis, gradable adjectives denote relations between individuals and degrees (Seuren 1973, Cresswell 1976 a.o). A gradable predicate, such as *tall*, incorporates the measure function *height*, which when applied to an individual, yields the degree *d* of *height* of that individual.

(6) $[\![\text{tall}]\!] = \lambda d\, \lambda x.\ \textbf{height}\ (x) \geq d$

In the degree analysis of adjectives, functional morphology such as, measure phrases (*two feet*), positive morphemes (POS), or the comparative morpheme *more* saturate the degree argument. Thus, the semantic composition for (7) is shown below:

(7) $[\![\text{Mary is six feet tall}]\!] = [\![\text{six feet}]\!]\ ([\![\text{tall}]\!])\ ([\![\text{Mary}]\!])$
$= \lambda x.\ \textbf{height}\ (x) \geq \textit{six feet}$

= **height** (*Mary)* ≥ *six feet*
'Mary's height is greater than or equal to six feet.'

This reading, where the measure phrase saturates the degree argument of the adjective and combines with the positive form of the adjective is called absolute measurement. In cases where the measure phrases combine with the comparative form of the adjective, the measure phrase provides the difference or the gap between the height of John and the height of Peter (as in (2a)), and the measure phrase is a predicate that gives the size of the gap (*one meter*, in ex. (2a)). This reading is referred to as the differential measurement reading.

In Spanish, the differential measurement reading follows Schwarzschild's (2005) interval-based semantics[3] where the measure phrase predicates over the interval from John's height to Peter's height in example (2a). The measure phrase gives the size of that interval: 'the interval is one meter'. Eguren and Pastor (2014) extend this analysis to measure phrases combining with bare adjectives in Spanish. They take the measure phrases in the [MP + *de* + A] construction to be predicative nominal expressions that are predicated of the adjective's degree argument, which is interpreted as an interval. In example (5), the gradable predicate *alta* 'high-FEM' undergoes Schwarzschild's homonym rule, which produces a homonym of the gradable predicate having an interval argument instead of a degree argument.

(8) a. ⟦dos metros de alta⟧ = Homonym rule in (8b) creates $alta_2$
Size of the interval I for $alta_2$ = dos metros

b. Homonym Rule: from degrees to intervals.
if A has meaning A' that relates individuals to degrees then A has a secondary meaning relating individuals to sets of degrees (intervals).
The secondary meaning is given by: λI. λx. I = {d: A'(x,d) }
(Schwarzschild, 2005:216)

c. ⟦Una valla de piedra de [dos metros de alta]⟧ = a fence that is high to an interval I, where I is two meters.

The theoretical analyses of measure phrases in Spanish, summarized above, makes the following predictions:

a) Measure phrases are possible with bare adjectives that belong to the lower closed scale.
b) Measure phrases can combine with the prepositional copula *de* and then combine with a gradable non-comparative predicate.
c) All measure phrases can combine with a comparative adjective.

Our study, using the acceptability rating methodology, tested these predictions on two groups of Spanish-English bilinguals, HS and L2 learners of Spanish. We turn to the study presently.

3 Acceptability and Usability Rating Study of Measure Phrases in Spanish

As described in the above section, the existing theoretical literature on Spanish measure phrases makes certain concrete testable predictions. We tested participants on both an acceptability rating and a usability rating task administered online to understand better what Spanish-English bilinguals would do with respect to preferences and use of measure phrases.

3.1 *Our Study* A total of 24 undergraduate students from Wichita State University, in Kansas, U.S.A., participated (6M, 18F). They were later divided into two groups on the basis of their answers to a background questionnaire: second language (L2) learners of Spanish and heritage speakers (HS). The HS had either immigrated to the U.S. from Mexico at age 6 or before, or they were born in the U.S. to immigrant parents; most had family origin in states of Northern Mexico (e.g. Chihuahua and Zacatecas).

[3] There are different analyses to the semantics of adjectives in Spanish. Due to space restrictions, we are unable to provide a summary. See García-Pardo and Menon (to appear) for an overview.

There were 13 L2 learners (4M, 9F) and 11 HS (2M, 9F). At the time of data collection, all of the participants were enrolled in intermediate-level Spanish language courses. Because this was a pilot study, the HS were not grouped according to whether they were simultaneous or sequential bilinguals, nor the amount of Spanish and English to which they were exposed as children.

3.2 *Methodology* We created an online acceptability rating study using a 2 * 2 design for the acceptability rating stimuli. The first group consisted of measure phrases combining with bare gradable adjectives. These adjectives were either comparative adjectives or non-comparative adjectives. The second group was the presence of absence of the prepositional copula *de*. Some examples are shown in the following table:

	Gradable Adjective	**Prepositional Copula**
Comparative	un metro **más** alto 'one meter taller'	
Non-comparative	# un metro alto[4] 'one meter tall'	
With *de*		dos pies **de** ancha 'two feet wide'
Without *de*		#dos pies ancha 'two feet wide'

Table 2: Sample experimental stimuli used in the acceptability rating study

3.3 *Design and Procedure* The questionnaire was administered on Qualtrics. Participants first read the university approved consent form. Once they gave their consent to participate in the survey, they provided socio-demographic information such as birthplace, age of arrival in the U.S., first language learned, etc. Following this, participants went on to rating combinations of measure phrases with comparative adjectives. Each participant rated 10 phrases. These phrases belonged to one of our four conditions. We chose 20 adjectives, a selection of these were taken from corpus data (Eguren and Pastor, 2014). Participants rated these phrases on a scale of 1-5, with 1 being highly unacceptable, 3 being neutral, and 5 being highly acceptable. They were also asked to answer, for each of the phrases, whether they use this phrase (or phrases like these) in their everyday speech. They were given three categories, 'natural for use in casual conversation', 'something that some people would use, but you wouldn't, and 'something that only a nonnative speaker would use'. These categories, roughly based on Labov (2006), allowed us to analyze the participants' perceptions of how the structure is used without confusing such perceptions with prescriptivist notions of what is considered 'correct' or 'incorrect.' The next page consisted of ten additional sentences with the prepositional copula *de*. No adjective was repeated between these two conditions. They rated these sentences on a scale of 1-5, and answered whether they use this sentence in their everyday speech. Following this, they proceeded to rate sentences with non-comparative adjectives and sentences without *de*. The procedure remained the same. Each participant rated 40 sentences in total for both acceptability and usability. The entire survey lasted 10-15 minutes. Participants belonged to intermediate and advanced Spanish grammar courses or conversation courses and were provided a link to take the survey either during class time or in their free time.

3.4 *Hypotheses* Given Montrul's (2010) findings that both HS and L2 learners exhibit cross-linguistic influence from English at the syntax-semantics interface, we hypothesized that we would find some similarities to English in both groups. For L2 learners, our hypothesis was that the majority of participants would allow measure phrases with any adjective, as they are learning the language and interference from their first language is possible. For HS speakers, we hypothesized that some would allow measure phrases with non-comparative adjectives or without *de* as they have learned those structures at home; however, we expected some of them to exhibit influence from English because, as research shows, the syntax-semantics interface seems to be permeable for HS as well.

[4] The # represents unacceptability and not ungrammaticality.

4 Results

Recall that we tested the participants on 4 different conditions. The first condition is called the comparative adjective condition (*un metro más alto* 'one meter taller'), the next condition is called the prepositional *de* condition (*dos pies de ancha* 'two feet wide'), followed by the non-comparative adjective condition (*#un metro alto* 'one meter tall') and without the prepositional *de* condition (*#dos pies ancha* 'two feet wide'). According to the theoretical literature, the non-comparative adjective condition and the without the prepositional *de* condition are disallowed generally.

First, we provide the mean ratings of acceptability provided by the L2 group and then the HS group. In the L2 group, participants rated the comparative adjective group a mean of 3.12 out of 5, and the non-comparative adjective group 2.87, thereby preferring adjectives in the comparative form to their non-comparative counterparts. The participants in the L2 group further rated the adjectives with the prepositional copula *de* a mean of 2.76 and adjectives without *de* a mean of 3.38, suggesting that they preferred measure phrases to combine with adjectives without the prepositional copula. The participants in the HS group rated the comparative adjective group a mean of 2.68, and the non-comparative adjective group 2.69. We find comparable ratings to both these conditions. In the HS group, participants also rated the adjectives with the prepositional copula *de* a mean of 3.52 and adjectives without *de* a mean of 2.99, suggesting that they prefer adjectives to combine with measure phrases using the prepositional copula. These results have been summarized in Table 3 and Figure 1 below:

	L2 group	**HS group**
Comparative *un metro más alto* 'one meter taller'	3.12	2.68
Non-comparative *#un metro alto* 'one meter tall'	2.87	2.69
With *de* *dos pies de ancha* 'two feet wide'	2.76	3.52
Without *de* *#dos pies ancha* 'two feet wide'	3.38	2.99

Table 3: Mean ratings of acceptability rating task

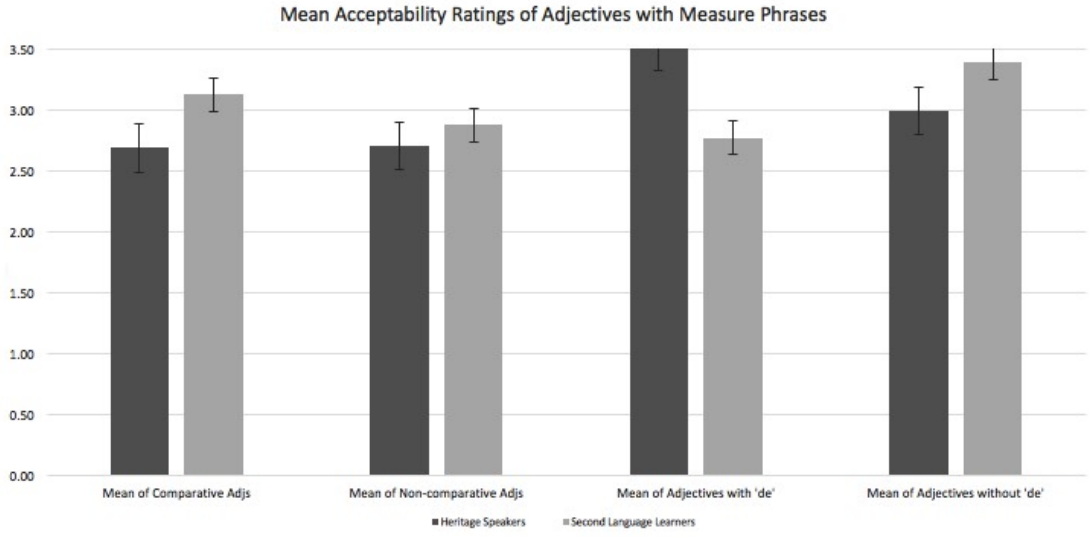


Figure 1: Mean acceptability ratings of measure phrases with adjectives. The darker columns are results of HS participants and lighter columns are results of L2 participants.

L2 learners preferred comparative adjectives to non-comparative adjectives, resembling the predictions for Spanish and contradicting our hypothesis; however, they also preferred adjectives without *de* to adjectives with *de*, which contradicts the normative Spanish usage. In fact, they preferred adjectives without *de* the highest of the two groups at 3.38. The latter is in line with our hypothesis, since this could be interference from their first language. The HS participants showed no preference between measure phrases with comparative adjectives and with non-comparative adjectives. They rated both these conditions exactly the same. This is in line with our hypothesis; their results are mixed, supporting our prediction that some HS would show influence from English while others would not. In addition, the HS group preferred adjectives with *de* the highest at 3.52. This has been summarized below:

	Comparative Adjectives *un metro más alto* 'one meter taller'	**Non-comparative Adjectives** *#un metro alto* 'one meter tall'
HS	No difference	No difference
L2	Preference	

Table 4: Summary of results of comparative adjectives and non-comparative adjectives

	Adjectives with *de* *dos pies de ancha* 'two feet wide'	**Adjectives without *de*** *#dos pies ancha* 'two feet wide'
HS	Preference	
L2		Preference

Table 5: Summary of results of adjectives with *de* and without *de*

Thus, we see that our HS participants did appear to exhibit influence from English, as hypothesized before, whereas our L2 learners preferred comparative adjectives, contrary to our hypothesis.

Next, we discuss the ratings of usability. Participants were asked to answer for each of the phrases whether they use this phrase (or phrases like these) in their everyday speech. They were given three categories, 'natural for use in casual conversation' (rating of 1), 'something that some people would use, but you wouldn't (rating of 2), and 'something that only a nonnative speaker would use' (rating of 3). They rated the same sentences as they did for the acceptability task. First, we provide the mean ratings provided by the L2 group and then the HS group.

The L2 group rated their usability of the comparative adjective group a mean of 1.74 out of 3, and the non-comparative adjective group 1.75 out of 3. The L2 group rated the adjectives with the prepositional copula *de* a mean of 1.83 and adjectives without *de* a mean of 2 out of 3. The HS group rated the comparative adjective group a mean of 2.05 out of 3, and the non-comparative adjective group 2.4 out of 3. The HS group rated the adjectives with the prepositional copula *de* a mean of 1.9 out of 3 and adjectives without *de* a mean of 2.53 out of 3. These results have been summarized in Table 6 below:

	L2 group	**HS group**
Comparative *un metro más alto* 'one meter taller'	1.73	2.05
Non-comparative *#un metro alto* 'one meter tall'	1.75	2.4
With *de* *dos pies de ancha* 'two feet wide'	1.83	1.9
Without *de* *#dos pies ancha* 'two feet wide'	2	2.53

Table 6: Mean ratings of usability rating task

As seen in the table, the L2 group rated the adjectives with measure phrases without the prepositional copula *de* the lowest, i.e. 2 (something that some people would use, but you wouldn't). The L2 group did

not distinguish between the usage of comparative adjectives versus non-comparative adjectives. They rated adjectives with *de* slightly better than adjectives without *de*. Overall, this suggests that L2 speakers use either measure phrases with comparative or non-comparative forms of the adjective, or measure phrases combining with the adjective using *de*, but they don't prefer to use adjectives combining with measure phrases without *de*. Similarly, the HS group also rated the use of adjectives without *de* the lowest. Similar to the L2 group, the HS group shows a comparable usability rating between adjectives in the comparative form and adjectives in the non-comparative form. They also preferred measure phrases combining with adjectives using the prepositional copula *de* the best. Thus, we see that both the groups have very similar usability ratings of the sentences. This is summarized in Figure 2 below.

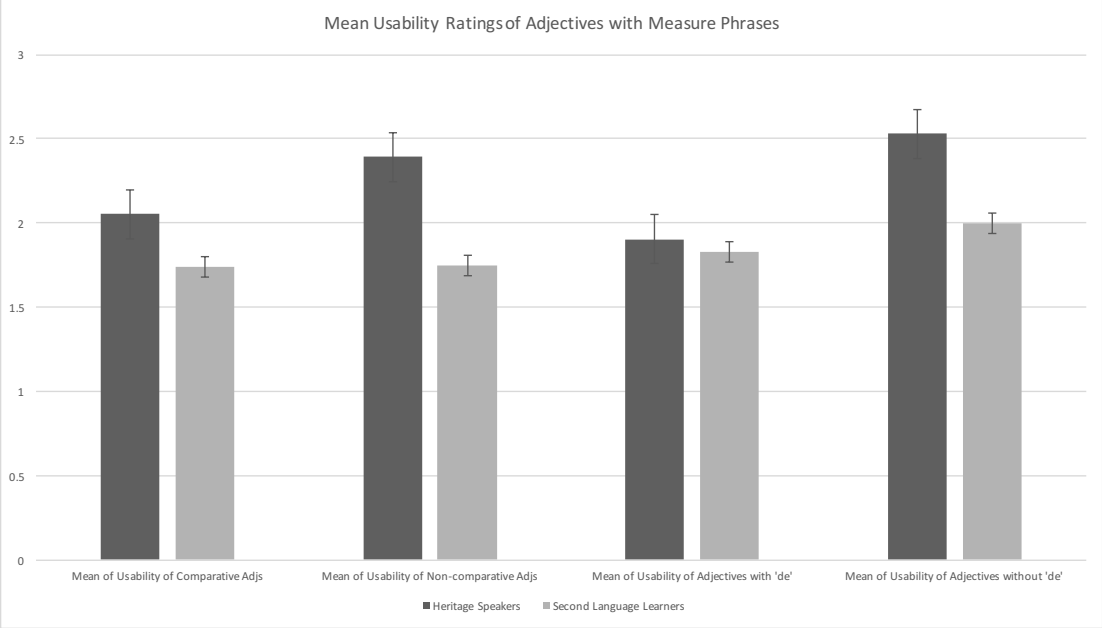


Figure 2: Mean usability ratings of measure phrases with adjectives. The darker columns are results of HS participants and lighter columns are results of L2 participants.

To summarize, both groups rated adjectives without *de* as the phrases they will least use in everyday speech. Both groups also preferred to use adjectives with *de* the most. Our results thereby provide new, experimental evidence for Eguren and Pastor's (2014) observation from corpus data, namely in Spanish measure phrases can combine with adjectives using the prepositional copula *de.*

	Comparative Adjectives *un metro más alto* 'one meter taller'	**Non-comparative Adjectives** #*un metro alto* 'one meter tall'
HS	Preference	
L2	No difference	No difference

Table 7: Summary of usability results of comparative adjectives and non-comparative adjectives

	Adjectives with *de* *dos pies de ancha* 'two feet wide'	**Adjectives without *de*** #*dos pies ancha* 'two feet wide'
HS	Preference	
L2	Preference	

Table 8: Summary of usability results of adjectives with *de* and without *de*

We find that the usability ratings differ from the acceptability ratings among participants and groups.

Recall that we hypothesized L2 learners could allow measure phrases with any adjective and some HS speakers should allow measure phrases with non-comparative adjectives and adjectives without *de*. Contra to our hypotheses, both participants showed a preference for adjectives with *de* and with measure phrases combining with comparative adjectives, thus suggesting that there is a difference between the language they use daily versus the language they find acceptable.

5 General Discussion

Our findings are in line with Montrul's (2010) observation that the HS in her study tended to display more influence from English at the syntax/semantics interface: our HS participants exhibited similarity to English in their preference for the type of adjectives that could be used with measure phrases, while they followed predictions for Spanish in their preference for the use of *de* with adjectives, a feature that is primarily syntactic. Montrul (2010) argues that the difficulty that some HS (particularly sequential bilinguals and early L2 learners) encounter at the syntax/semantics interface may be an indication that they have not reached 'ultimate attainment' of the language. The notion that a certain group of HS has not reached 'ultimate attainment' or has achieved an 'incomplete acquisition' of their heritage language suggests that there is an inherent deficiency in the linguistic repertoires of these speakers, and therefore has received a great deal of scrutiny from linguists (e.g. Otheguy and Zentella, 2012). Following Pascual y Cabo and Rothman (2012), we prefer not to use the term 'incomplete' to describe the grammatical systems of HS who have acquired English at a young age and do not utilize all of the grammatical structures that appear in the variety of Spanish spoken by their parents; rather, we simply describe their systems as 'different.' One possible explanation for the HSs' lack of alignment with predictions for Spanish is that they may possess a 'simplified' grammatical system due to early acquisition of English; however, the reasons for differences between HS and L2 grammars are still not fully understood (Bayram et al., 2016). Further information on the age of first exposure to English and exposure to Spanish in educational contexts would be helpful to determine whether participants' input in Spanish was limited from a young age and whether they had opportunities to develop their academic literacy in the language. As Bayram et al. (2016) point out, recent research shows that HS who receive significant literacy training in the HL as part of their primary education show minimal differences from monolinguals of the same age as early adults (Bayram et. al. cite Kupisch & Rothman, 2016).

Another possible explanation for the HSs' preference for both comparative and non-comparative adjectives is that they may have acquired a contact variety of Spanish either at home or in bilingual contexts outside of home (Escobar and Potowski, 2015). For example, perhaps in some communities of speakers of U.S. Mexican Spanish (e.g. a network of families from Zacatecas who live in Kansas), it could be acceptable to use measure phrases with non-comparative adjectives; the existence of such a feature could be analyzed as the result of structural borrowing from English. In this case, the HS selection of measure phrase + non-comparative adjective combinations on the questionnaire would be analyzed as a contact feature rather than a result of a process of language acquisition (e.g. transfer, generalization, etc.). In order to determine whether this is the case, it would be necessary to collect additional data from other groups of speakers of U.S. Mexican Spanish in the same community, including the parents of the study participants.

While we have presented two feasible explanations for the HSs' apparent indifference to the comparative vs. non-comparative adjective distinction, there is also a third possibility: their responses may have been related to the study design rather than their linguistic awareness. Because the phrases the participants judged were presented out of context and were not complete sentences, it is possible that they may have judged some of the phrases that were intended to represent normative Spanish usage as unacceptable for reasons other than the measure phrase/adjective structure. For example, having learned Spanish in a natural setting at home, they may have been unaccustomed to examining short written phrases out of context, as L2 learners may have been.[5] It would be helpful to utilize additional complementary

[5] This difference in linguistic background also explains why L2 classes are often difficult for heritage speakers, especially when taught with a focus on traditional mechanical grammar exercises.

forms of data collection to further examine this feature, such as a production task. In particular, our results from the usability task suggests that there is a difference between rating a phrase versus actually using the phrase. A production study would shed more light on this difference.

Regarding the L2 learners' preference for comparative adjectives, this finding may be a reflection of the level of proficiency in Spanish that had been reached by the students who participated in this study. To be allowed in the intermediate Spanish classes where data were collected, students had to have either completed a three-semester basic language sequence or scored high enough on a placement exam to skip ahead. Many of the students have had some study abroad or missionary/public service experience in Spanish-speaking countries by this point in their studies. However, using the students' Spanish class level as an indication of language proficiency may not provide enough information to compare HS with L2 learners; a more precise method of measuring students' language proficiency would help to support such conclusions. Moreover, it may be more useful to compare highly proficient L2 learners with HS (rather than students classified as 'intermediate') as both groups have been classified as having reached an 'end-state' (Bayram et al., 2016).

Our examination of the syntax and semantics of measure phrases for HS and L2 learners expands our understanding of the grammars of these groups, contributing to the fields of cognitive linguistics, sociolinguistics, and the research base for heritage language pedagogy (Bayram et al., 2016), as well as complementing existing theoretical studies in syntax and semantics. It is important for heritage language instructors to understand the differences between how some HS may express measure phrases and the ways such phrases would be structured by Spanish monolinguals, and to critically reflect on their understanding of these differences, in order to teach HS better in the language classroom. Instructors' understanding of HS syntax/semantics can support contemporary critical and sociolinguistic approaches to heritage language education (e.g. Leeman, 2005; Shin and Henderson, 2017) by providing additional information about the reasons for the differences between the grammars of different groups of Spanish speakers. Further research is needed to better understand this aspect of Spanish HS syntax/semantics and how heritage language instructors should address it in the classroom.

6 Future Work

The authors intend to collect additional data using both a perception and a production task and testing a broader range of participants, including different proficiency levels and, for the HS, different ages of first exposure to English. In addition to HS and L2 learners, participants will also include first generation Spanish speakers from northern Mexico who live in central Kansas, including the parents of the HS participants. We hope to be able to include first generation Spanish speakers with both high and low levels of proficiency in English, and we also hope to be able to use a standardized method to evaluate the Spanish proficiency level of all of the participants.

References

Bayram, Fatih, Prada, Josh, Pascual y Cabo, Diego, & Rothman, Jason. 2016. Why should formal linguistic approaches to heritage language education be linked to heritage language pedagogies? *Handbook of Research and Practice in Heritage Langauge Education*: Springer International Publishing AG.

Bosque, Ignacio. 1998. Sobre los complementos de medida. In *Estudios en honor del profesor Josse de Kock*, eds. Nicole Delbecque and Christian De Paepe, 57-73. Leuven: Leuven University Press.

Cresswell, Max J. 1976. The semantics of degree. In Barbara Partee (ed.) Montague Grammar. New York: Academic Press, 261-292.

Eguren, Luis and Pastor, Alberto. 2014. Meausre phrases with bare adjectives in Spanish. *Natural Language and Linguistics Theory* 32: 459-497.

Escobar, Anna María, & Potowski, Kim. 2015. *El español de los Estados Unidos*. Cambridge: Cambridge University Press.

García, Ofelia. 2009. *Bilingual education in the 21st century: A global perspective*. Malden, MA and Oxford: Wiley/Blackwell.

García-Pardo, Alfredo and Menon, Mythili. to appear. The Aspectual Structure of the Adjective: Spanish *Ser* and *Estar*. *Issues in Hispanic and Lusophone Linguistics*. John Benjamins Publishing.

Kennedy, Chris, and McNally, Louise. 2005. Scale structure and the semantic typology of gradable predicates. *Language*, 81(2), 345–381.

Kupisch, Tanja, & Rothman, Jason. 2016. Terminology matters! Wy difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*, 1-19. doi:10.1177/1367006916654355

Labov, William. 2006. *The social stratification of English in New York City* (2nd Edition). Cambridge: Cambridge University Press.

Leeman, Jennifer. 2005. Engaging critical pedagogy: Spanish for native speakers. *Foreign Language Annals, 38*(1), 35-45.

Montrul, Silvina. 2010. Dominant language transfer in adult second language leaners and heritage speakers. *Second Language Research, 26*(3), 293-327.

Otheguy, Ricardo, & Zentella, Ana Celia. 2012. *Spanish in New York: Language contact, dialectical levelling, and structural continuity*. Oxford: Oxford University Press.

Pascual y Cabo, Diego, & Rothman, Jason. 2012. The (Il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics, 33*, 450-455.

Sáez del Álamo, Luis. 1997. Cuantifacadores y sintagma de medida. In *Proceedings of the IV colloquium of generative grammar*, eds. Natàlia Català and María Bargalló, 162-184. Tarragona Universitat Rovira i Virgili.

Sawada, Osamu and Grano, Thomas. 2011. Scale structure, coercion, and the interpretation of measure phrases in Japanese. *Natural Language Semantics*, 191-226.

Schwarzschild, Roger. 2005. Measure phrases as modifiers of adjectives. *Recherches linguistiques de Vincennes* 34: 207-228.

Seuren, Pieter. 1973. The Comparative. In Ferene Kiefer and Nicolas Ruwet (eds.) Generative Grammar in Europe. Dordrecht D. Reidel. 528-564.

Shin, Naomi, and Henderson, Mary. 2017. A sociolinguistic approach to teaching Spanish grammatical structures. Foreign Language Annals, 50:1, 195-213.

Showstack, Rachel. 2017. La pragmática transcultural de los hablantes de herencia de español: análisis e implicaciones pedagógicas. *Journal of Spanish Language Teaching, 3*(2).

Silva-Corvalán, Carmen. 1993. On the permeability of grammars: Evidence from Spanish and English contact. In William J. Ashby, Marian Mithun, and Giorgio Perissinotto (eds.), *Linguistic Perspectives on Romance Languages: Selected Papers from the XXI Linguistic Symposium on Romance Languages, Santa Barbara, February 21–24, 1991* (pp. 19-43): Benjamins.

Toledo, Assaf, and Sassoon, Galit. 2011. Absolute vs. relative adjectives – Variance within vs. between individuals. In *Proceedings of SALT 21*, 135-154.

Interaction of Merge and Labeling: Consequences for Hyperraising

Manabu Mizuguchi
Toyo University

1 Introduction

Hyperraising is one instance of movement in which the *n*P goes to Spec-T across a finite clause with tense and agreement. Consider examples from (1) through (3):

(1) *John seems [that t is sick][1]

(2) Babaandu ba-lolekhana [mbo t ba-kwa][2] Lubukusu
2people 2SA-seem that 2SA.PST-fall
'The people seem like they fell/The people seem to have fallen.' (Carstens & Diercks 2013)

(3) Eng'ombe chi-bonekhana [t chi-ng'were amachi] Lussamia
10cow 10SA-appear 10-SA-drink 6water
'The cows appear to have drunk the water.' (Carstens & Diercks 2013)

As illustrated, hyperraising is not possible in English while it is in Bantu languages, showing cross-linguistic differences. In this paper, I take Lubukusu and Lussamia as representatives of Bantu languages. The argument that (2) and (3) are examples of hyperraising, and not those of copy-raising, is evidenced by the fact that the surface subject can consistently be interpreted as being under the scope of the matrix predicate and that deduction interpretation is allowed (Carstens & Diercks 2013). Copy-raising, on the other hand, does not allow this reading. Consider (4), which is felicitous only as an observation about John or his appearance; it cannot be stated as a deduction when one finds that John is absent (Rogers 1974, Potsdam & Runner 2001):

(4) John seems like he is sick.

In addition to cross-linguistic variations, what is interesting about hyperraising is that it is not always possible in Bantu. As shown in (5) and (6), unlike in (2) and (3), the *n*P cannot hyperraise when different complementizers are selected in the embedded clause:

(5) *Mikaeli a-lolekhana [a-li t a-si-kona] Lubukusu
Michael 1SA-seem that 1SA-PERS-sleep
'Michael seems to be still sleeping.' (Carstens & Diercks 2013)

(6) Eng'ombe chi-bonekhana [koti chi-ng'were amachi] Lussamia
10cow 10SA-appear that 10SA-drink 6water
'The cows appear as if they have drunk the water.' (Carstens & Diercks 2013)

In (6), reconstructed reading of *Eng'ombe* is not at all possible, suggesting that copy-raising, not hyperraising, is behind the derivation of (6).

* I would like to thank the audience at WECOL 2017 for comments and questions. The research reported here was supported by Japan Society for the Promotion of Science under Grant-in-Aid for Young Scientists (B) (#16K16835).

[1] "t" and angle brackets "〈 〉" indicate copies and are used interchangeably.

[2] The abbreviations used in this paper are as follows: SA = subject agreement, PST = past tense, PERS = persistive. Arabic numerals preceding nouns in glosses of Bantu examples indicate noun class.

Proceedings of *WECOL 2017*

Given the empirical facts, the following questions will arise: When is hyperraising allowed and what mechanisms are behind it? In this paper, I address these questions and discuss hyperraising cross-linguistically and intra-linguistically. I claim that hyperraising is deduced from Merge and labeling, the elements of an explanatory theory of language, and is given a principled explanation.

This paper is organized as follows. In Section 2, I discuss a framework of principled explanation in quest of an explanatory theory of language, spelling out the basic hypothesis. In Section 3, I argue that hyperraising in English and Bantu is explained by interaction of Merge and Label. In Section 4, I show that the analysis proposed for hyperraising also explains the *that*-trace effect. In Section 5, I demonstrate that Merge accounts for lack of the EPP effect. In Section 6, I summarize and conclude the paper.

2 The Strong Minimalist Thesis

In this section, I first discuss a framework of principled explanation in linguistics. The basic assumption in an explanatory theory of language is to seek the simplest account of Universal Grammar (UG), which is reasonable for scientific and biological reasons given that linguistics is a branch of science and that language is a biological endowment: scientifically, simpler theory suggests deeper explanation and leads to better understanding of the object of inquiry; in any scientific inquiry, less is better than more. Biologically, simplification of UG, which is the theory of the faculty of language, makes possible an eventual account of the origin of language as far as this can be attained (Chomsky 2015). In the best case, linguistic phenomena would be explained by Merge, a virtual conceptual necessity, and the interfaces, through which narrow syntax interacts with the external systems so that language can be put to use. This hypothesis is articulated by the Strong Minimalist Thesis (SMT) (Chomsky 2000 et seq., especially Chomsky 2010):

(7) SMT: Merge + interfaces = language

Applying iteratively to form syntactic objects (SOs) or sets, Merge interacts with general principles of minimal computation or third factor principles, and applies in the simplest possible way. Constrained by such general principles, Merge is formulated as simplest Merge and yields a set out of any two SOs:[3]

(8) Merge(α, β) = {α, β}

At the interfaces, Full Interpretation must be satisfied; otherwise, interface conditions will be violated. The bare minimum requirement for Full Interpretation is identification of SOs created by simplest Merge, which are label-less and hence, unidentified. Conceptual-Intentional (CI) and externalization processes must know what kind of object a derived SO is and SOs must be identified for interpretation (identifiable = interpretable at the interfaces). Chomsky (2013, 2015) proposes that identification of SOs is carried out by a labeling algorithm called Label. The algorithm, which is understood as nothing more than one instantiation of minimal search, determines the properties, hence "labels," of SOs. When applied to the set marked as α in (9), Label qua minimal search locates within α the closest head (in the case of a set of the form {X, {YP}} as in (9a)) or agreeing heads (in the case of an {{XP}, {YP}} set as in (9b)), providing necessary information about it for its CI interpretation and externalization, hence for Full Interpretation at the interfaces:

(9) a. [$_\alpha$ X [Y …]] (α labeled X)
b. [$_\alpha$ [X_χ …] [Y_χ …]] (α labeled <χ, χ>)

Labeling by Label is one necessary component of Full Interpretation and is motivated by the interfaces. As far as language is explained by Merge and Label, the simplest, hence principled, account of language will be achieved, and language will be perfectly designed to satisfy interface conditions.

With this background in place, I now consider hyperraising in English and Bantu, demonstrating that its grammatical behaviors are explained by interaction of Merge and Label.

[3] Curly brackets and square brackets are used interchangeably ({α, β} = [α, β]). In this paper, Greek letters are used for general symbols and for set labels. For ease of distinction, those used as general symbols are shown in italics.

3 Hyperraising in English and Bantu

3.1 *Ill-formed Hyperraising* Let us start with the discussion of English hyperraising (=(1)). I assume a phase-based derivational model proposed by Chomsky (2007, 2008), where operations except External Merge (EM) apply at the phase level for proper interpretation at the interfaces. This means that the movement or Internal Merge (IM) of the embedded subject will apply when CP or the κ-marked set is structured in the derivation as in (10) through the merge of C, a phase head:

(10) [κ that [λ T [α be [John sick]]]]

Given (10), two derivational scenarios are possible for hyperraising. In the first scenario (Scenario 1), the *n*P *John* moves to Spec-T. The movement, however, will baffle hyperraising. As discussed in Epstein, Kitahara & Seely (2012) and in Mizuguchi (2014), counter-cyclic A-movement will induce Transfer, which allows the IM to apply without causing a failure of the derivation. Epstein, Kitahara & Seely (2012) argue that since Merge abides by the No-Tampering Condition (NTC), A-movement creates a double-peaked structure (or an intersecting set) (=(11)), one peak of which (i.e., γ, which includes λ) must be eliminated through Transfer for the derivation to continue; a double-peaked structure, they say, is inaccessible to further Merge, which halts the derivation. On the other hand, in Mizuguchi (2014), I claim that transfer of λ makes the derivation of A-movement possible by allowing the *n*P to be merged at the root. In (12), λ is embedded and is not a root, with the result that IM of the *n*P with λ, as it infixes the SO, will tamper with the existing structure, violating NTC. If λ is transferred, however, the *n*P can be internally merged at the root, not violating NTC:

(11) [κ C
[λ T [α ⟨John⟩ […]]]
[γ John
⇒ Transfer

(12) [C [λ T [α John […]]]]
⇓ Transfer
[γ John [λ T [α ⟨John⟩ […]]]]
(order irrelevant)

Whether we adopt Epstein, Kitahara & Seely (2012) or Mizuguchi (2014), they both argue that given simplest Merge, A-movement of the *n*P will necessarily lead to transfer of γ or λ; otherwise, the derivation will fail. If γ or λ is transferred upon A-movement, then the subject will become syntactically invisible and cannot be subject to further computation in the higher clause as it is transferred as part of γ / λ. The *n*P cannot raise out and the derivation of hyperraising halts in the embedded clause. Scenario 1 derivationally rules out hyperraising.

Now consider the other scenario. In Scenario 2, as shown in (13), the *n*P moves to Spec-C directly, which allows it to avoid being transferred as the movement is cyclic, applying at the root. Hence, it can hyperraise:

(13) [σ John [κ that [λ T [α be [⟨John⟩ sick]]]]]

Under this scenario, however, λ will not be labeled and labeling failure will result. In English, T alone is too weak to label and for it to work as a label, Spec-T must be visible when Label applies (Chomsky 2015). In (13), there is no overt element in Spec-T upon the application of the labeling algorithm and λ will not be labeled. Under Scenario 2, hyperraising is ruled out at the interfaces in violation of Full Interpretation.

In summary, Merge and Label exclude hyperraising under the two derivational scenarios.

Now consider hyperraising in Bantu. Ill-formed hyperraising is ruled out in the same way as English hyperraising: the movement out of γ fails for Transfer under Scenario 1 and labeling fails for weak T under Scenario 2. Moreover, ϕ-features on T will be under-valued under Scenario 2. It has been argued that in Bantu, ϕ-feature agreement is executed locally via Spec-head (Kinyalolo 1991, Ndayiragije 1999, among others) or in the XP-YP configuration under the current theoretical framework, where Spec is used just for convenience.[4] The Bantu examples in (5) and (6) are explained on par with the English example in (1).

[4] This argument also applies to English if Agree is reducible to Label, being executed as part of XP-YP labeling, as proposed, say, in Abe (2016), Fukui (2017), Kitahara (2017) and Seely (2016).

3.2 *Proposal: Grammatical Hyperraising* If hyperraising is ruled out either derivationally or at the interfaces as I have discussed, then why is the movement possible in (2) and (3)? I argue that grammatical hyperraising is explained by Merge and propose the third scenario (Scenario 3) under SMT. Under the assumption of simplest Merge, Merge, operating freely, applies asymmetrically as well as symmetrically, and two modes of Merge are available: that is, set-Merge, which produces a single set out of two SOs as in (14a) and pair-Merge, which adjoins one SO to the other and yields an ordered pair as in (14b):

(14) a. $\{\alpha, \beta\}$ b. $<\alpha, \beta>$

Given simplest Merge, these two modes of Merge are available without any assumptions or stipulations.[5]

With (14) in mind, I claim that in (2) and (3), C and T are not set-merged as in (15), which produces (10) in the embedded clause. In (15), T is set-merged with α, a set headed by *v*/*v** and then C is set-merged with λ, a set headed by T:

(15) a. $[_{\lambda}$ T $[_{\alpha} \ldots]]$ b. $[_{\kappa}$ C $[_{\lambda}$ T $[_{\alpha} \ldots]]]$

Instead, in the relevant examples, T is externally pair-merged to C and then the pair-merged SO (i.e., a head-head amalgam) is set-merged with α. Consider the derivation in (16):[6]

(16) a. <C, T> b. $[_{\delta}$ <C, T> $[_{\alpha} \ldots]]$

Under Scenario 3, (16b) is produced in the embedded clause. In the derivational process, in order to raise out, hence hyperraise, into the higher clause, the *n*P is internally merged with the δ-marked set, which yields (17) (here I illustrate the derivation with the Lubukusu example):

(17) $[_{\mu}$ babaandu $[_{\delta}$ <C, T> $[_{\alpha}$ t [ba-kwa]]]]

The proposed theory of hyperraising can solve the problems posed by Scenarios 1 and 2. First, the *n*P *babaandu* is not transferred via its movement to the Spec of <C, T> since the δ-marked set is not embedded and the movement can apply at the root; it does not induce transfer of μ or δ. Second, λ or the T-headed set is not formed in the derivation thanks to pair-merge of T to C. Hence, there is no need to worry about labeling of TP. Finally, since T is not set-merged and is not available in the derivation, feature-inheritance to T does not take place and C retains ϕ (and tense). Agreement can be executed locally via Spec-head or XP-YP, when the *n*P is internally merged with δ to form μ, and ϕ will not be under-valued.[7] None of the problems we face under Scenarios 1 and 2 (i.e., a transfer problem posed by Scenario 1 and labeling and under-valuation problems posed by Scenario 2) arises under Scenario 3.

The derivation continues from (17) to produce (18), which yields (2):

(18) [babaandu [ba-lolekhana $[_{\mu}$ t $[_{\delta}$ <C, T> $[_{\alpha}$ t [ba-kwa]]]]]] (⇒ (2))

Notice that the movement from the Spec of <C, T>, which yields a copy in the Spec, does not cause labeling problems and hence that the derivation in (18) is not at all problematic, either. SOs, when pair-merged to others, get de-activated and become syntactically invisible as adjoined SOs are asymmetrical to their hosts (Chomsky 2004, 2015, Mizuguchi 2017);[8] <C, T>, as T is adjoined to C, retains all the properties of C and

[5] As noted in Fukui (2017), the ordered pair $<\alpha, \beta>$ is mathematically equivalent to the set $\{\alpha, \{\alpha, \beta\}\}$ (or $\{\{\alpha\}, \{\alpha, \beta\}\}$ in the Kuratowski definition). If so, (14b) can be derived by symmetrical Merge or set-Merge and there is no need to have both set-Merge and pair-Merge, with Merge applying only symmetrically. For expository convenience, I use set-Merge and pair-Merge for (14a) and (14b), respectively.

[6] Given simplest Merge, pair-Merge, like set-Merge, applies both externally and internally. For relevant discussion, see Epstein, Kitahara & Seely (2016). In this paper, unless otherwise mentioned, (pair-)Merge is external (pair-)Merge.

[7] Notice that given the proposed theory, "under-inheritance" (Legate 2011) receives a Merge-based, hence principled, explanation: under-inheritance of ϕ-features follows from pair-merge of T to C.

[8] De-activation or invisibility of β for its pair-merge to α or for β being asymmetrical to α is attributable to self-

is syntactically on par with C. Since C, unlike T, is strong as a label, it can label without overt Spec. Labeling problems do not arise with μ and δ even if the *n*P moves out and becomes a copy, which is invisible to Label.

The final question is why Scenario 3 does not apply to (5) and (6). Merge applies freely to the extent that it abides by third factor principles and only a stipulation can block pair-merge of T to C in the relevant examples. I argue that the way C and T are merged results in different phonological realizations of complementizers, proposing that <C, T> and C are externalized as distinct complementizers. Consider (19), which summarizes syntax-externalization relationships of complementizers in the two Bantu languages discussed in this paper:

(19) COMPLEMENTIZER REALIZATIONS

Syntax		Externalization	
<C, T>	→	mbo	Ø
C	→	a-li[9]	koti
(Languages :		Lubukusu	Lussamia)

The two kinds of complementizers in Bantu reflect how Merge applies to C and T. If *a-li* and *koti* indicate that C is merged as an independent head, then T is not pair-merged to C and Scenario 3 does not take place in the derivation of (5) and (6).

Summarizing the discussion in this section, I have claimed that hyperraising in English and Bantu is explained by interaction of Merge and Label. Since these two operations are ingredients of SMT, hyperraising receives a principled explanation under the theory proposed here.

4 <C, T> in English: The *That*-Trace Effect

In the last section, I have proposed that the way C and T are merged leads to different externalizations of complementizers in Bantu languages. Given that the proposal is correct, it is expected that it also holds true of other languages. I argue that it does. I claim that the zero complementizer Ø in English is an externalization of C with T pair-merged to it; on the other hand, when T is not pair-merged to C, C is realized as the overt complementizer *that*. Consider (20), which updates (19) and summarizes different realizations of complementizers in the languages discussed in this paper:

(20) COMPLEMENTIZER REALIZATIONS IN BANTU AND ENGLISH

Syntax		Externalization		
<C, T>	→	mbo	Ø	Ø
C	→	a-li	koti	that
(Languages :		Lubukusu	Lussamia	English)

I submit that English is like Lussamia in the way C is externalized.[10]

With this proposal in place, I show that the *that*-trace effect, one instance of the ECP phenomena, falls under the proposed theory. The *that*-trace effect is typically illustrated by the following examples:

(21) a. *Which student do you think that read the book?
b. *Ted is one of those who George believes that will win the prize.

(22) a. Which student do you think read the book?
b. Ted is one of those who George believes will win the prize.

embedding by α. Recall from footnote 5 that $<\alpha, \beta>$ is mathematically on par with $\{\alpha, \{\alpha, \beta\}\}$. Since β is embedded in α, it cannot probe out of $\{\alpha, \{\alpha, \beta\}\}$; nor can it be located by minimal search as α is closer. Hence, computation by β and targeting β both fail (i.e., β is syntactically de-activated).

[9] More precisely, *a-li* is an agreeing complementier and *-li* is the complementizer; *-li* agrees with the most local subject in the higher clause. For details, see Diercks (2013).

[10] See Pesetsky & Torrego (2001), who claim that the overt complementizer *that* is not C but originally T; for them, *that* is a spell-out of T that has head-moved to C (i.e., *that* is <C, T>) and the analysis, they argue, explains the *that*-trace effect. As I discuss immediately below, the assumption that *that* is T is not necessary to account for the ECP phenomenon.

Ā-movement of the subject out of the embedded clause is possible when the complementizer is not present. I argue that grammatical behaviors of (21) and (22) are explained in the same way as Bantu hyperraising. Let us start with (21), taking (21a) for illustration. In (21), the overt complementizer indicates that C and T are set-merged (=(15)), with the λ-marked set produced in the derivation of the embedded clause. Consider (23):

(23) [κ C [λ T [α which student [*v** …]]]]

Given (23), the subject *wh*-phrase will move either to Spec-T (Scenario 1) or to Spec-C (Scenario 2) in the next step in order to move out. If the *n*P moves to Spec-T, then it will be trapped in the embedded clause and cannot move out: it is transferred as part of γ and becomes invisible, hence inaccessible, to further computation due to counter-cyclic movement of the *n*P. As discussed in Section 3.1, the derivation of the embedded clause will fail when counter-cyclic A-movement is executed, unless Transfer applies to γ or λ:

(24) [✗ [κ C [γ which student [λ T [α t [read the book]]]]]]
↓
that

Extraction of the *n*P is derivationally ruled out by counter-cyclic A-movement, hence by Merge.

On the other hand, if syntax chooses Scenario 2 and the *n*P moves to Spec-C, skipping Spec-T, it can evade being transferred and can be subject to further computation as the movement is not counter-cyclic. In this case, however, λ will not be labeled in the absence of overt Spec-T: T is too weak to label λ and Label cannot identify λ though T is located by the operation through minimal search:

(25) [τ which student [κ C [λ T [α t [read the book]]]]]
↓ ✗ labeling: λ = unlabeled for weak T
that

The derivation will be ruled out at the interfaces for labeling failure.

Now consider the examples in (22), where subject extraction out of the embedded clause is well-formed. I take (22a) for illustration. When the overt complementizer is not present, T is pair-merged to C and the amalgamated head <C, T> is produced in syntax. (26), instead of (23), is yielded in the embedded clause:

(26) [δ <C, T> [α which student [*v** …]]]

In the next step, the *wh*-phrase moves to the Spec of <C, T> to get out of δ. The movement does not trigger transfer of δ or μ in (27a), hence the subject, as it applies at the root and is not counter-cyclic; *wh*-movement does not halt in the embedded clause. Furthermore, since λ or the T-headed set is not formed in the derivation thanks to T being pair-merged to C, there is no need to worry about labeling of TP and a labeling problem does not arise at all. Recall that the movement of the subject out of μ does not raise labeling problems with μ and δ, either: <C, T> can work as a label without overt Spec since C alone can label. As illustrated in (27), (22a) can be derived without causing transfer and labeling problems:

(27) a. [μ which student [δ <C, T> [α t [read the book]]]]
↓ ✔transfer; ✔labeling
Ø
b. [which student [C [… [t [*v** [… [μ t [δ <C, T> [α t [read the book]]]]]]]]]]
✔labeling of μ and δ

In sum, the grammatical contrast between (21) and (22) is explained by interaction of Merge and Label. As I have demonstrated, the *that*-trace effect and hyperraising are explained in the same way by the theory

proposed in this paper. The two empirical phenomena receive a principled account in terms of SMT.

If the overt complementizer *that* is an indication that T is set-merged, with the T-headed set or λ produced in the derivation, Scenario 3 cannot be employed for examples like (1), either. Consider once again (1), which shows that hyperraising is not possible in English. The example is repeated below for convenience as (28):

(28) *John seems [that t is sick] (=(1))

The example, where the overt complementizer is present, is derived only under Scenario 1 or 2 as T is not pair-merged to C. As I have discussed in the last section, under both these scenarios, the example causes problems and is ruled out either derivationally for counter-cyclic A-movement, which will induce Transfer and halt hyperraising, or at the interfaces for labeling failure, which is due to weak T.

Under the theory proposed in this paper, the *that*-trace effect is explained by SMT and assumptions beyond Merge and Label are not at all necessary to account for the grammatical contrast between (21) and (22). Hence, it can be concluded that the proposed theory is principled. In the rest of this section, I discuss Chomsky's (2015) recent analysis of (21) and (22), and argue that the assumptions he introduces are all additional assumptions beyond SMT and that given the theory proposed here, they can be dispensed with.

Chomsky (2015) introduces the following assumptions to explain subject extraction in (21) and (22):[11]

(29) a. complementizer deletion (C → Ø)
b. phasehood inheritance
c. phase-level memory

Chomsky argues that in examples such as (22), where the subject can be successfully extracted, C, a phase head, and along with it, the C-headed set, are deleted and that with it, phasehood is inherited by T together with all the other properties of C (e.g., ϕ and tense). These processes, according to Chomsky, allow γ and λ to be labeled <ϕ, ϕ> and T respectively, through A-movement of the subject to Spec-T in γ and still enable the subject to move out of the embedded clause (= γ) in the next phase without de-labeling γ and λ even if a copy of the subject is created in the embedded Spec-T by its movement:

(30) [ε which student do you think [Ø [γ t [λ T [α t [*v**-read the book]]]]]]
✔γ and λ = labeled

Chomsky (2015), unlike Chomsky (2007, 2008), assumes that not only EM but also other operations can apply at the non-phase level; A-movement to Spec-T occurs cyclically, before C, a phase head, is merged. Thanks to (29c), copies can be distinguished from repetitions even if IM does not apply at the phase level.

On the other hand, when *that* is not deleted as in (21), the embedded C keeps its phasehood and movement of the subject out of the embedded Spec-T will de-label γ and λ. Since a copy of the subject in Spec-T is invisible to Label, labeling failure will result with γ and λ in (31) for weakness of T as a label:

(31) [ε which student do you think [κ that [γ t [λ T [α t [*v**-read the book]]]]]]
✗γ and λ = unlabeled

The derivation violates Full Interpretation for unidentified SOs and rules out (21) at the interfaces.

Complementizer deletion (29a) and phasehood inheritance (29b) make labeling of γ and λ possible even when the subject moves out in the next phase and the embedded Spec-T becomes a copy for its movement. Chomsky argues that this is due to phase-level memory (29c): γ and λ stay labeled up to the next higher phase (ε in (30) and κ in (31)). In (30), γ and λ have been labeled before *that* deletes and the label information is available in ε thanks to inheritance of phasehood to T; ε is now a single phase including γ. Consequently, γ and λ are kept labeled and re-labeling is not required in ε. In (31), on the other hand, the relevant information is only available in κ, the lower phase, and is lost in ε, the higher phase; γ and λ, which have been de-labeled,

[11] Pesetsky (2017) puts forward a similar proposal, arguing that the *that*-trace effect is due to deletion or "exfoliation" of the C-layer. See section 6 for some discussion.

have to be re-labeled, which fails due to an invisible copy in Spec-T.

Chomsky justifies the assumptions in (29) by saying that the observation of ECP as in (21) is normal and its violation by the zero complementizer or *that*-deletion as in (22) is an idiosyncratic case. Notice, however, that the assumptions are all additional assumptions beyond SMT (i.e., Merge and the interfaces) and hence are theoretically unfavorable. Specific problems of (29) are as follows: (29a), which is crucial to Chomsky, violates the No-Tampering Condition (NTC), which requires that SOs produced in syntax be left unchanged. This is because the embedded C and along with it, the C-headed set {C, γ}, are deleted and the existing structure is tampered with. Given that syntax is constrained by third factor principles, deletion of structures is not allowed and cannot be executed.

As for (29b), phasehood, under the current phase theory, is an intrinsic property of specific heads; in other words, phasehood is not a feature. If C deletes, then the phasehood should simply disappear and phasehood will not be inherited. Moreover, phasehood inheritance raises a question of why phasehood, unlike ϕ and tense, is inherited only when C deletes. Chomsky (2013, 2015) argues that features on non-phase heads are all inherited from phase heads. In the case of C, both of its ϕ and tense are inherited onto T. If phasehood is inherited just like ϕ and tense, then why is phasehood not inherited when ϕ and tense on C are inherited? Why does phasehood inheritance obtain only via C deletion? A reasonable conclusion is that phasehood inheritance is not like feature inheritance, which further strengthens the argument that (29b) is just stipulated.

Finally, as for (29c), phase-level memory, which has been proposed as an alternative to indexing, is not motivated anywhere else and is assumed only to warrant label information in (30), hence to explain the *that*-trace effect. As I have noted above, Chomsky (2015:11) argues that phase-level memory is also required to distinguish copies from repetitions, allowing phase-internal operations to apply at the non-phase level without violating interface conditions. As Chomsky (2007, 2008) argues, however, phase-internal operations such as IM and valuation can be executed in an interface-compliant manner if they apply at the phase level, where syntax interacts with the interfaces via Transfer and relevant information (say, EM or IM, syntactically valued or not) is made locally available to the interfaces (or to the interface components). The assumption of phase-level applications of operations or interaction of syntax with the interfaces through Transfer (see also Mizuguchi 2016), which is inevitable for CI and SM interpretations, makes phase-level memory redundant.

Under the theory proposed in this paper, the grammatical contrast between (21) and (22) follows from SMT and nothing more needs to be assumed; the assumptions in (29) can be wiped out. To be more specific, the contrast is reducible to whether pair-Merge applies, hence to Merge. If T is pair-merged to C, the γ-marked and λ-marked sets will not be produced and the subject *wh*-phrase can move out of the embedded clause without causing transfer and labeling problems. Moreover, the violation of ECP in (22) is not at all idiosyncratic: it is deducible from Merge, a virtual conceptual necessity, under the proposed theory.[12]

5 Pair-Merge and lack of the EPP

In the discussion of grammatical hyperraising in Bantu and English examples in (22), where the subject can be extracted when *that* is not present, the crucial ingredient is pair-Merge: in the derivation, T is pair-merged to C (=(32a)) instead of T being set-merged (=(32b)):

(32) a. $[_{\delta} \langle C, T \rangle [_{\alpha} \dots]]$ b. $[_{\kappa} C [_{\lambda} T [_{\alpha} \dots]]]$

(32a) allows the derivation to proceed without causing any problems: when the relevant pair-merge applies, λ or the T-headed set will not be yielded in the derivation, which solves transfer and labeling problems.

In this section, I discuss one consequence of this mode of Merge. Chomsky (2013, 2015) argues that the EPP effect is reducible to labeling: T by its nature is too weak to label and in order to work as a label, it requires overt Spec-T (i.e., some overt SO must be merged with λ) when Label applies. If pair-merge of T to C solves the labeling problem as it does not yield λ, then the EPP effect will not be observed when the relevant pair-merge applies. I argue that lack of the EPP effect is the result of pair-merge of T to C. I consider Icelandic and Yiddish to show that this argument is correct and that Merge explains the EPP effect.

[12] Nomura (2015) argues, contrary to what I have proposed, that the grammatical contrast between (21) and (22) is explained by affixation of C to T by internal pair-Merge. See also Pesetsky & Torrego (2001) for the discussion of the *that*-trace effect (footnote 10).

Let us start with Icelandic. In this language, Verb Second (V2) is observed. As examples, consider (33):

(33) a. Það hafa verið seldir margir bílar á þessu uppboði. Icelandic
there have been sold many cars at this auction
'Many cars have been sold at this auction.' (Jonas 1996)
b. Það hafa komið margir menn hingað í dag.
there have come many men here today
'Many men have come here today.' (Koeneman & Neeleman 2001)

V2 in Icelandic applies at the C level: that is, SOs are moved to Spec-C and the verb is moved to C (Vikner 1995 among others). The expletive *það* 'there' appears only in V2 contexts when nothing is moved to Spec-C for V2. It is thus merged in Spec-C, not in Spec-T. This is evidenced by (34), where the expletive is ruled out when some SO (*hér* in (34a) and *í gær* in (34b)) is moved to Spec-C for V2:

(34) a. Hún sagði að hér drykkju (*það) margir vín. Icelandic
she said that here drink (there) many [people] wine
'She said that here many people drink wine.'
b. Í gær var (*það) mikill snjór á jörðinni.
yesterday was (there) much snow on ground-the
'Yesterday there was much snow on the ground.' (Maling & Zaenen 1978)

In (33), the *n*Ps *margir bílar* and *margir menn* come after the main verbs *seldir* and *komið* and it can be considered that they stay in the positions where they are externally merged and do not move to Spec-T. Given that the expletive is in Spec-C, Icelandic thus shows lack of the EPP effect.

Lack of the EPP effect follows from pair-Merge. It has been standardly assumed that in V2, verb movement obtains and that the verb, along with T, goes all the way up to C; in V2, these heads are derivationally amalgamated. Such amalgamation is executed by pair-Merge in the Merge-based theory of syntax assumed here (Chomsky 2015). Suppose that the verb is not head-moved (i.e., internally pair-merged) to T and then to C as commonly assumed in the literature; instead, amalgamation of heads in V2 is done by both external and internal applications of pair-Merge, which are allowed under the simplest-Merge hypothesis: in the derivation of V2, T is externally pair-merged to C and the verb is internally pair-merged to the amalgamated head; it is head-moved to <C, T>, not to T. The derived structure is (35):[13]

(35) $[_{\mu}$ Spec $[_{\delta}$ <[<C, T>], V> [… t_V …]]]

In (35), T is affixed, hence adjoined, to C by external pair-Merge, and λ or the T-headed set is not formed in the derivation. Thus, labeling failure will not result even if the *n*P is not internally merged.

The EPP effect is also absent in (36), where the *n*P *margir bílar* moves to the clause-initial position:[14]

(36) Margir bílar hafa verið seldir á þessu uppboði.

Since the expletive is not merged in (36), *margir bílar* moves to Spec-C and not to Spec-T. The proposed derivation is also behind (36) and explains lack of the EPP effect.[15]

Now suppose that T is not pair-merged to C but is set merged as in (32b). In this case, $\{_{\lambda}$ T, {*v*P}} is formed in the derivation and merge of the *n*P with λ, hence movement to Spec-T, will be required to label

[13] Notice that the verb is also an amalgamated head since *v* and V are pair-merged when the verb head-moves to *v*. For ease of exposition, the pair-merge is abstracted away in (35). Likewise, it does not matter here whether V is pair-merged to *v* or the pair-merge applies in the other way around, and how it applies (externally or internally).

[14] See, e.g., Craenenbroeck & Haegeman (2007), Schwartz & Vikner (1996), Vikner (1995) for the argument that the subject-initial clause is also a V2 clause.

[15] Notice that in (33) and (36), the *n*P cannot be merged in Spec-T and then internally merged to form Spec-C: if so, a copy of the *n*P will be created in Spec-T by IM, which is invisible to Label, and the sets γ and λ in $\{_{\gamma}$ ⟨*n*P⟩, $\{_{\lambda}$ T, {*v*P}}} will not be labeled as T is too weak to label without overt Spec.

the set; the EPP effect will be observed. Consider (37):

(37) [$_{\tau}$ Spec [$_{\kappa}$ C [$_{\gamma}$ *n*P [$_{\lambda}$ T [$_{\alpha}$ … t …]]]]] (λ = labeled)

This derivation is endorsed by (38), where the *n*P moves to the position after the tensed verb in C and before the participle, which stays in its externally merged position:[16]

(38) a. Það hafa margir bílar verið seldir á þessu uppboði.
b. Það hafa margir menn komið hingað í dag.

The position in question can be considered Spec-T as the expletive is in Spec-C.

Yiddish is another language which demonstrates the absence of the EPP effect. It has been observed that the EPP effect disappears in the embedded clause when the verb and the subject are inverted (=(39)); otherwise, the EPP effect is observed (=(40)). To see this, consider the following examples:

(39) Vos hot er nit gevolt [zoln [mir leyenen t]]? Yiddish
what has he not wanted should we read
'What did he not want us to read?' (Diesing 1990)

(40) a. Vos hot er nit gevolt [$_{\kappa}$ az [$_{\gamma}$ mir [$_{\lambda}$ zoln [leyenen t]]]]? Yiddish
what has he not wanted that we should read
b. ?Vos hot er nit gevolt [$_{\kappa}$ az [$_{\gamma}$ es [$_{\lambda}$ zoln [mir leyenen t]]]]?
what has he not wanted that EXPL should we read (Diesing 1990)

Diesing (1990:72) argues that the Spec-T position can remain empty, hence the EPP effect does not appear, when the complementizer is not present, in which case the verb moves to C to govern the empty Spec-T. As evidenced by (41), if the overt complementizer *az* is present, which blocks verb movement, the subject cannot remain in its first-merged position (Spec-*v**) and Spec-T must be formed:

(41) *Vos hot er nit gevolt [$_{\kappa}$ az [$_{\lambda}$ zoln [$_{\alpha}$ mir leyenen t]]]?

Lack of the EPP effect in (39), and hence Diesing's observation, is explained straightforwardly if pair-Merge is assumed. I argue that when the complementizer is present, T is set-merged and the T-headed set is yielded in the derivation; in other words, the overt complementizer is an externalization of C as in Lusamia and in English (C → *az*). (41) is ruled out as ill-formed for λ being unlabeled: overt Spec-T is not formed and T cannot work as a label when located by minimal search Label.

The zero complementizer, on the other hand, is an externalization of C with T pair-merged to it (<C, T> → Ø). In this case, since the T-headed set is not produced, movement to Spec-T is not required for labeling of the set; the EPP effect is thus not observed. In Yiddish, which is also a verb-movement language, the verb combines with tense via head movement and the verb is internally pair-merged to <C, T>, as in Icelandic (=(35)). This yields (39). Diesing's observation follows from pair-merger of T to C.

To summarize, I have discussed Icelandic and Yiddish and have shown that pair-merge of T to C explains lack of the EPP effect in a principled way; the presence or absence of EPP follows from Merge. Notice that the proposed theory can explain the generalization that V2 languages allow the optionality of EPP: V2 is generated by free applications of Merge (i.e., by external as well as internal pair-merge of T to C).

[16] In (38), where the verb is in C, T is not externally but internally pair-merged to C, with (i) derived from (37) (verb movement to T and verb movement to *v* are abstracted away for ease of exposition):

(i) [$_{\tau}$ Spec [$_{\kappa}$ <C, T> [$_{\gamma}$ *n*P [$_{\lambda}$ ⟨T⟩ [$_{\alpha}$ … t …]]]]]

The pair-merge yields a copy of T, which is visible to Label (Epstein, Kitahara & Seely 2016). A labeling problem does not arise with λ and γ, however: in (i), overt Spec-T is created by merge of the *n*P with λ and the sets can be identified by Label as T strengthens by overt Spec-T and can serve as a label.

6 Conclusion

In this paper, assuming SMT as a framework of principled explanation, I have claimed that hyperraising in English and Bantu is deduced from interaction of Merge and Label. Hyperraising thus receives a principled explanation and is one illustration of the hypothesis that language is an optimal solution to interface conditions. I have also argued that *that*-trace and EPP phenomena follow from the proposed theory, eliminating additional assumptions beyond Merge and Label.

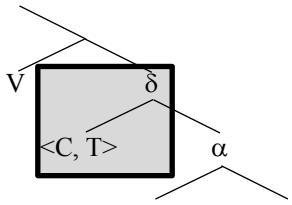
The discussion in this paper supports Pesetsky's (2017) generalization (42), which says that clause size matters for extraction of the subject:

(42) Subject extraction always entails a smaller-than-full clause.

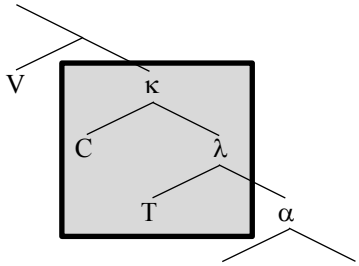
Pesetsky discusses hyperraising and the *that*-trace effect as instances that fall under (42). As noted in footnote 11, he proposes Exfoliation to explain (42): thanks to Exfoliation, a certain portion of a clause (say, CP) is deleted or removed, which yields a reduced clause and makes subject extraction possible.[17]

Under the theory proposed in this paper, clause reduction is not due to deletion or exfoliation but is attributed to pair-merge of T to C, hence to Merge; clause size follows from the way Merge applies:

(43) Clause reduction by pair-Merge



(44) Full-clause production by set-Merge



As shown in (43), external pair-Merge can reduce structure and the proposed analysis can give a principled explanation to clause reduction; no extra mechanism beyond Merge (such as literal deletion operations like *that*-deletion and Exfoliation, which tamper with the existing structure and violate NTC) is assumed to explain clause reduction. Moreover, interaction of Merge and Label explains why clause reduction is required for subject extraction: unless the clause is reduced, the subject will be trapped in the embedded clause for Transfer due to its counter-cyclic movement and cannot move out or labeling failure will arise with λ for weak T. Clause reduction for subject extraction is explained by SMT and Pesetsky's empirical generalization is deduced in a principled manner under the theory I proposed in this paper.

I conclude that the simplest account of UG is achieved by the proposed theory of hyperraising: language keeps only to the two irreducible, hence principled elements hypothesized in SMT: Merge, the simplest recursive operation, and the interfaces (i.e., labeling for Full Interpretation in this particular case).

References

Abe, Jun. 2016. Dynamic antisymmetry for labeling. *Lingua* 174:1–15.

Carstens, Vicki & Michael Diecks. 2013. Parameterizing case and activity: Hyper-raising in Bantu. In *NELS 40*, ed. by Seda Kan, Claire Moore-Cantwell & Robert Staubs, 99–118. Amherst: University of Massachusetts, Graduate Linguistics Student Association.

Chomsky, Noam. 1981. *Lectures on government and binding: The Pisa lectures*. Dordrecht: Foris.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. by Roger Martin, David Michaels & Juan Uriagereka, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In *The cartography of syntactic structures*. Vol. 3, *Structures and*

[17] The generalization and a rule of deletion to account for it are not new; they go back to Chomsky (1981), where S′-deletion (i.e., deletion of CP) is proposed to explain subject extraction.

beyond, ed. by Adriana Belletti, 104–131. Oxford: Oxford University Press.
Chomsky, Noam. 2007. Approaching UG from below. In *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics*, ed. by Uli Sauerland & Hans-Martin Gärtner, 1–29. Berlin: Mouton de Gruyter.
Chomsky, Noam. 2008. On phases. In *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*, ed. by Robert Freidin, Carlos P. Otero & Maria Luisa Zubizarreta, 133–166. Cambridge, MA: MIT Press.
Chomsky, Noam. 2010. Some simple evo devo theses: How true might they be for language? In *The evolution of human language: Biolinguistic perspectives*, ed. by Richard K. Larson, Viviane Déprez & Hiroko Yamakido, 45–62. Cambridge: Cambridge University Press.
Chomsky, Noam. 2013. Problems of projection. *Lingua* 130:33–49.
Chomsky, Noam. 2015. Problems of projection: Extensions. In *Structures, strategies, and beyond: Studies in honour of Adriana Belletti*, ed. by Elisa Di Domenico, Cornelia Hamann & Simona Matteini, 1–16. Amsterdam: John Benjamins.
Craenenbroeck, Jeroen & Liliane Haegeman. 2007. The derivation of subject-initial V2. *Linguistic Inquiry* 38:167–178.
Diercks, Michael. 2013. Indirect agree in Lubukusu complementizer agreement. *Natural Language and Linguistic Theory* 31:357–407.
Diesing, Molly. 1990. Verb movement and the subject position in Yiddish. *Natural Language and Linguistic Theory* 8:41–79.
Epstein, Samuel David, Hisatsugu Kitahara & T. Daniel Seely. 2012. Structure building that can't be. In *Ways of structure building*, ed. by Myriam Uribe-Etxebarria & Vidal Valmala, 253–270. Oxford: Oxford University Press.
Epstein, Samuel David, Hisatsugu Kitahara & T. Daniel Seely. 2016. Phase cancellation by external pair-merge of heads. *The Linguistic Review* 33:87–102.
Fukui, Naoki. 2017. *Merge in the mind-brain: Essays on theoretical linguistics and the neuroscience of language*. New York: Routledge.
Jonas, Dianne. 1996. Clause structure and verb syntax in Scandinavian and English. Doctoral dissertation, Harvard University, Cambridge, MA.
Kinyalolo, Kasangati Kikuni Wabongambilu. 1991. Syntactic dependencies and the spec-head agreement hypothesis in KiLega. Doctoral dissertation, UCLA, Los Angeles, CA.
Kitahara, Hisatsugu. 2017. Maximizing syntactic minimization: Five preliminary studies. *Reports of the Keio Institute of Cultural and Linguistic Studies* 48:239–265.
Koeneman, Olaf & Ad Neeleman. 2001. Predication, verb movement and the distribution of expletives. *Lingua* 111:189–233.
Legate, Julie. 2011. Under-inheritance. Paper presented at the 42nd North East Linguistic Society, University of Toronto, 12 November.
Maling, Joan & Annie Zaenen. 1978. The nonuniversality of a surface filter. *Linguistic Inquiry* 9:475–497.
Mizuguchi, Manabu. 2014. Phases and counter-cyclicity of A-movement. In *Proceedings of the 16th Seoul international conference on generative grammar*, ed. by Jong Un Park & Il-Jae Lee, 257–277. Seoul: The Korean Generative Grammar Circle.
Mizuguchi, Manabu. 2016. De-activation of ϕ through pair-Merge. *English Linguistics* 33:108–118.
Mizuguchi, Manabu. 2017. Labelability and interpretability. *Studies in Generative Grammar* 27:327–365.
Ndayiragije, Juvénal. 1999. Checking economy. *Linguistic Inquiry* 30:399–444.
Nomura, Masashi. 2015. Pair-Merge and its consequences. Paper presented at the 87th General Meeting of the English Literary Society of Japan, Rissho University, 23 May.
Pesetsky, David. 2017. Exfoliation: Towards a derivational theory of clause size. Paper presented at Workshop on quirks of subject extraction, National University of Singapore, 11 August.
Pesetsky, David & Esther Torrego. 2001. T-to-C movement: Causes and consequences. In *Ken Hale: A life in language*, ed. by Michael Kenstowicz, 355–426. Cambridge, MA: MIT Press.
Potsdam, Eric & Jeffrey Runner. 2001. Richard returns: Copy raising and its implications. In *CLS 37*, ed. by Mary Andronis, Christopher Ball, Heidi Elston & Sylvain Neuvel, 453–468. Chicago: The Chicago Linguistic Society.
Rogers, Andy. 1974. A trans-derivational constraint on Richard? In *CLS* 10, ed. by Micheal W. La Galy, Robert A. Fox & Anthony Bruck, 551–558. Chicago: The Chicago Linguistic Society.
Seely, T. Daniel. 2016. Is Agree labeling? Paper presented at Keio Linguistics Colloquium, 6 June.
Schwartz, Bonnie D. & Sten Vikner. 1996. The verb always leaves IP in V2 clauses. In *Parameters and functional heads: Essays in comparative syntax*, ed. by Adriana Belletti & Luigi Rizzi, 11–62. Oxford: Oxford University Press.
Vikner, Sten. 1995. *Verb movement and expletive subjects in the Germanic languages.* Oxford: Oxford University Press.

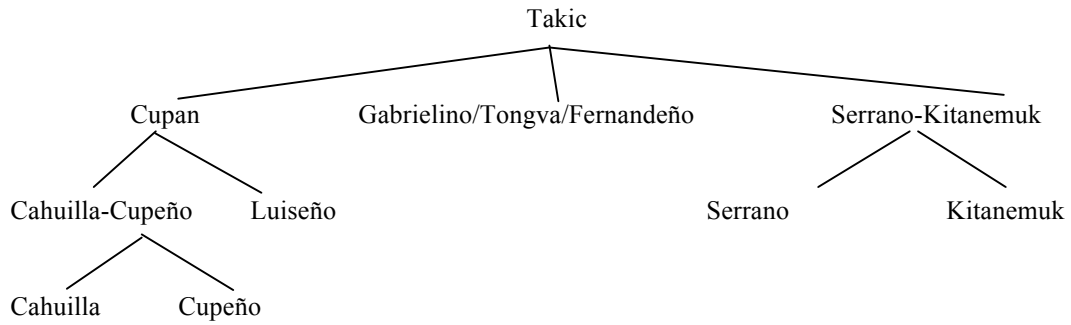
The Tongva Series II Pronominal Clitics Once Again

Pamela Munro
UCLA

1 Background

1.1 *The language* Gabrielino/Tongva/Fernandeño (henceforth Tongva)[1] is one of the least known languages of the Uto-Aztecan family, a member of the Takic subfamily. Formerly spoken throughout the Los Angeles basin, Tongva has probably not been spoken by native speakers for over 60 years. All primary documentation of this language comes from work by linguists and others during the 19th and early 20th centuries. The largest documentation, and that I've worked with most extensively, is in the notes of J. P. Harrington (henceforth JPH), but there are additional sources of earlier data, many of which are carefully transcribed and collected in McCawley (1996). These earlier vocabularies (especially those from C. Hart Merriam [1903]; henceforth CHM) are less reliable phonetically, but they contain many important items not found in JPH's corpus. Unfortunately, there are many gaps in our knowledge of both grammar and lexicon. Secondary documentation is in Munro (in progress a, b).

Below is what I believe about the interrelationships among the Takic languages:



(There were certainly other Takic languages, including Nicoleño (Munro with Johnson 2001) and Tataviam (Munro 2002),[2] which I will not consider here.) Currently it seems best to me to regard Takic as having three coordinate branches, as diagrammed above (following Kroeber 1907), although others have proposed linking Gabrielino with either Cupan (Bright 1974) or Serrano-Kitanemuk (Miller 1984).

* Thanks to Harrington and Merriam for their careful if sometimes eccentric notes, and to my fellow members of the Gabrielino-Tongva Language Committee (especially Brent Scarcliff and Allen Munro), as well as to the other Uto-Aztecanists who have discussed the Tongva language with me, including the late William Bright, the late Alice Anderton, the late Donald Crook, Geraldine Anderson, Jane Hill, Kenneth Hill, and Marcus Smith. I'm also grateful for comments when I presented this paper during the 2017 Friends of Uto-Aztecan Conference meeting held along with WECOL in Boise.

[1] Traditionally, the language I'll focus on here has been called Gabrielino by linguists. We have very little data on the language known to linguists as Fernandeño, but I will assume that it is a closely related dialect and not consider it separately further here. The language is now most commonly called Tongva, Gabrielino-Tongva, or Tongva-Gabrielino by heritage learners and other interested groups south of the San Fernando Valley. "Gabrielino" is sometimes spelled "Gabrieleno", "Gabrieliño", or "Gabrieleño". There is also an outlier group favoring the name "Kizh" (Tongva *kiiy* 'house'), following Buschmann (1865). I use the name Tongva following the preferences of the Tongva members of the Gabrielino-Tongva Language Committee.

[2] Tatatviam is also the name now used by many modern Fernandeño-Tataviam descendants for the language (variety) known in the literature as Fernandeño, but I refer here only to a very poorly documented language of the northern San Fernando Valley.

1.2 *Tongva sounds and orthography* All data here is presented in our current Tongva orthography, which is based closely on JPH's presumed phonemicization. The Tongva alphabet, in alphabetical order is *', a, aa, b, ch e, ee, f, g, h, i, ii, k, kw, l, m, n, ng, o, oo, p, r, s, sh, t, u, uu, v, w, x, y, z*. In this practical orthography, *'* = [ʔ], *ng* = [ŋ], *ch* = [č], *sh* = [š], and *kw* = [k^w]; other letters have their expected phonetic values (though there is some indeterminacy about the exact phonetic value of the vowel written here as *o*). It is possible that some confusion of sounds may have arisen during various transcription processes: for example, some cases of *r*-*v* confusion appear to have occurred as notes were copied and recopied (including by JPH and his assistants).

It is not yet clear whether Tongva vowels are distinguished only for stress or for length, or for both stress and length; certainly, however, these two features are closely correlated, since there are very few unstressed long vowels in the corpus (except in some words with plural possessor prefixes such as *'eyoo-* 'our', which seem to have an unusual (proclitic?) status), and the great majority of stressed vowels are marked as long. If a stressed vowel is written long in any of JPH's recordings of a given word, we write it long (doubled); the first such doubled vowel in a word is stressed.[3] Vowel quality is correlated with stress: unstressed short /e/ and /i/ neutralize to /e/, and /o/ and /u/ to /o/.[4]

1.3 *A brief typological overview of Tongva* As in most Uto-Aztecan languages, Tongva word order is highly variable.

Verbs inflect for a few categories, but comparative evidence from other Takic languages suggests that many other verbal forms, especially for verbs of subordinate clauses, were not documented. The most common verb form in the documentation is what I call the BASIC VERB (small caps indicate terminology from Munro in progress a), which can receive almost any non-future indicative translation.

The language has nominative-accusative case marking and a few other oblique case markers on nouns. Noun plurality is indicated by reduplication or suffixation or both (many nouns have several documented plurals).

There is no case marking on independent pronouns, which are fairly rarely used; non-third-person pronominal arguments are shown with second position clitics. There are two sets or "series"[5] of subject-object clitics (Munro (2000) is a preliminary description).

The language also has mood clitics, including indicative (NORMAL) *='e*, imperative (COMMAND) *='aa* (singular) / *='avo* (plural), interrogative (QUESTION) *=ha'*, and subjunctive (WISH) *=po*. These clitics do not cooccur, but all but the last are often omitted in relevant examples. Their potential presence defines four moods.

2 The Tongva Series I clitics

Since word order in Tongva is very free, second-position clitics or clitic sequences can follow any sentence element. Sentences may include only a pronominal clitic, only a mood clitic, both,[6] or (with only third-person singular arguments in indicative sentences) neither.[7] JPH almost always wrote the clitic or clitic combination attached to an adjacent word in the sentence. Usually this was the preceding word (and this is how I write them), but

[3] Words with no long vowels for which stress is marked in the corpus are written with a long vowel in place of the stressed vowel (many of these, incidentally, have stressed diphthongs in closed syllables; the original recording is recoverable in our dictionary entries). There are only a few words in which neither stress nor length is marked (other than one-syllable words). Some roots lose stress with certain affixes (e.g. *kiiy* 'house' has forms like *'eyookinga* "in our house"): this phenomenon seems analogous to the "stressless root" phenomena described for the Cupan languages (Hill and Hill 1968).

[4] In earlier versions of the dictionary (Munro in progress b), only the only unstressed/short vowels used in the orthography were *e* and *o*. However, we have revised the system to write unstressed *i* and *u* in cases where stress shifts show the underlying quality of the stressed vowel. For example, 'blackberry' was formerly written *pekwaa-r*, but its reduplicated plural, *pi-piikwa-r*, shows that the singular should be written *pikwaa-r* (though it would not contrast phonetically with another word *pekwaa-r*). Newer versions of the dictionary thus employ a more morphological, less strictly phonological system that makes formation of plurals, possessed forms, and other derivatives more regular, less dependent on memorization.

[5] As noted in Munro (2000), employment of this term is a tribute to the late Wick Miller, who had used it to refer to Kawaiisu.

[6] The indicative clitic *='e* follows pronominal clitics; the other mood clitics precede them.

[7] In addition there is a mystery clitic *=m* (seen for example in the second sentence in (1) below; similarly described in Munro 2000), about which I can so far say very little.

sometimes the clitic or clitic complex was recorded attached to a following word. However, clitics never occur initially. The pronunciation of the clitic complex changes in some phonological environments.

The Series I clitics are shown in Chart 1 at the end of this handout, along with some simple examples. Here are three important points about their use alone and with mood particles:

• Clitics occur in second position. For Tongva, this means after the first word; crucially, though, if there is an initial phrase, such as a noun phrase, in the sentence, the Series I clitic follows that phrase.

• Clitics do not occur initially.

• Mood clitics are (in general) not obligatory, but they do not cooccur. The indicative mood clitic occurs after the pronominal clitics; the other mood clitics precede pronominal clitics.

3 The Tongva Series II clitics

3.1 *The Series II clitics* Munro (2000) presented the following list of the Series II clitics, suggesting that their final *y*'s are derived from the indicative mood clitic *=ʼe*:[8]

=ney / =noy / =nay	equivalent to Series I *=ne* (1s)
=rey / =roy / =ray	equivalent to Series I *=re* (1p and 1s > 2s)
=ʼay / =ʼey	equivalent to Series I *=ʼa* (2s)
=y	equivalent to Series I zero (3s)
=mey	equivalent to Series I *=me* (3p; also 3s > 2)

Further examination of the data shows that this is not a complete list. Additional variants of *=mey* include (as we might expect) *=may* and *=moy,* and there are many uses of the Series II clitics not listed above. One unusual use is "environmental" (maybe something like an ethical dative), where a first person argument (usually plural) appears in a statement about weather or the like.

We don't know what determines a speaker's choice to use a Series II clitic. Often sentences with both types of clitic are volunteered as synonymous: [9]

(1)	Kwa'yook=noy	ma'eete.	/	Noo=n='e=m[10]	kwa'yook	ma'eete.	
	be.afraid.B=*noy*	much		pron.1s=*n*=ind=?	be.afraid.B	much	'I am very scared'

In all the data below pronominal clitics are glossed with their surface forms (since almost all of these have more than one meaning). Series II clitics may be identified by their final *yʼ*s.

There are over 300 examples of this set of clitics in the data (including a number from CHM), but it's hard to count; many are repeated, and it's not always clear if they are re-elicitations or just additional copies of a single recording.

Below is a chart with examples of all the uses of these clitics that I've found, along with presumed Series I equivalents. Problematical examples (discussed below) are boldfaced, with problematical clitics boxed and unexpected Series I equivalents followed by !!!.

[8] JPH's brief comments on some data suggest he may have had the same idea.

[9] Abbreviations used in the glosses include abs = absolutive, acc = accusative, B = basic verb, fut = future, ind = indicative, loc = locative, P = posssessive, pl = plural, pron = pronoun, red = reduplication, quot = quotative, subj = subjunctive. Pronouns and pronominal prefixes are glossed with 1, 2, 3, s, p. Pronominal series I and II clitics are glossed with their surface form.

[10] This is the mystery clitic *=m* mentioned in footnote 8.

Chart 2: Tongva Series II Clitics and Their Uses

arguments	*Series II*	*examples*	*Series I* ?
1s intransitive	*=nay*	Too'avo-r=nay. 'I am blind'[11] blind-abs=*nay*	*=ne='e* (usually realized as *=n='e)*; with subj. *=po*, *=po=ne* !!!; never initial !!!
	=ney	Tehoovko=ney xaa. 'I am well' well=*ney* be **Xaay=po=ney xaa**. 'No, I'm not' (CHM)[12] not=subj=*ney* be	
	=noy	Karuukmo-ro=noy. 'I am going to stand up' stand-fut=*noy* Kwa'yook=noy ma'eete. 'I am very scared' be.afraid.B=*noy* much **Noy koviinok**. 'I am hungry' *noy* be.hungry.B	
1s > 3s transitive	*=nay*	'Ekwaa=nay woshaa'ax 'amooya'-a. 'I am here looking at the dead here=*nay* look.at.B dead-acc person' Worooy-t-a worooy-t-a=nay koo-ro. 'I will call to each man'[13] man-abs-acc man-abs-acc=*nay* call-fut	*=n='e, =na'*
	=ney	Yawiinok=ney. 'I believe him'[14] believe.B=*ney*	
	=noy	Twiinok=noy meyaa kotaa-ra. 'I am burning this stick' burn.B=*noy* this wood-acc	
3s > 1s transitive	*=ney*	Kokook=ney motuuche-y. 'A flea bit me' bite.B=*ney* flea-abs **Pe'ee worooy-t paare=ney**. 'That man said thus to me' that man-abs quot=*ney*	*=n='e*
	=noy	Chiipax=noy. '[The cat] scratched me' scratch.B=*noy* **Menee worooy-t mokaa-ro=noy**.[15] 'This man is going to kill me' this man-abs kill-fut=*noy*	

[11] The same sentence is given as *Too'avo-r=noy* by CHM. See section 3.2.

[12] This is the only subjunctive sentence in the Series II dataset (it's not completely clear how *=po* contributes to its meaning). This sentence was given as an answer to *Hyaa='a xaa?* 'Are you ready?' [now=*'a* be.B].

[13] I'm guessing this is the best translation of "Al hombre al hombre voy a llamar", based on a similar usage with time words (e.g. *tameevngey-a tameevngey-a* 'every year').

[14] JPH translated this sentence as "(yo) te creo" ('I believe you'), but for this we'd expect *Yawiinok=rey*. Mistake in either transcription or translation? This is the only 1s > 3s example for *=ney*.

[15] This sentence was given as equivalent to *Menee worooy-t=ney mokaa-ro* and *Mokaa-ro=ney menee worooy-t*, each with a second-position clitic.

arguments	*Series II*	*examples*	*Series I* ?
3s > 1s environmental	=*nay*	Xaxaaytok='e ne-shuun-nga tamaave-t, meyaa[16]=nay bleed.B=ind 1sP-heart-loc magic-abs this=*nay* pwiinok.[17] "Magic is bleeding in my heart, it is filling up." fill.up.B	=*n='e*
	=*ney*	Muuro=ney wakook. 'It rained on me there' there=*ney* rain.B	
2s > 1s imperative	=*ney*	**Maxaa=ney paarama'.** 'Give me that'[18] give=*ney* that	=*ne='aa* !!!
1p intransitive	=*rey*	Hyaa'mo=rey yaata-r-om. 'Now we are sleepy' now=*rey* sleepy-abs-pl	=*re='e* (usually realized as =*re=')*
	=*roy*	Yovaa-ng'aro=roy mii. 'We are going to the church' church-to=*roy* go.B	
1p > 3s transitive	=*roy*	Poochenax=roy. 'We are crushing it'[19] crush.B=*roy*	=*re=', =ra'*
1s > 2s transitive	=*rey*	Waak-ro=rey. 'I'm going to comb your hair' comb-fut=*rey*	=*re='*
	=*roy*	Totaa-rar=roy piik. / **Roy piik totaa-rar.** rock-with=*roy* throw.B 'I threw at you with a rock'	
3s > 1p environmental	=*ray*	Wakook=ray. 'It's raining now' (CHM) rain.B=*ray*	=*re='*
	=*rey*	'Ahiiken=rey kii. 'A wind is coming' wind=*rey* come.B	
	=*roy*	'Akwaaken=roy choohenax chaav-t-a. 'The rain put out the fire' rain=*roy* extinguish.B fire-abs-acc Muuro=roy wakook. 'It rained on us there' there=*roy* rain.B	
3s > 1s transitive	=*rey*	Kokoo[20]=rey. 'It bit me' bite=*rey*	=*n='e* !!!
	=*roy*	Muhuuk=roy. 'He shot me' (CHM) shoot.B=*roy*	
3s intransitive	=*rey*	Rawaate'=rey kii. 'A white man is coming' (CHM) white.man=*rey* come.B	=*'e* !!!
3s > 3s transitive[21]	=*rey*	Kwi'tii'=rey mokaa-ro. 'The boy wants to [will] kill it' boy=*rey* kill-fut	=*'e* !!!

[16] Tongva has many, many demonstrative roots whose distribution is not well understood. 'This' and 'that' are somewhat impressionistic glosses.
[17] Normally this seems to be an intransitive verb, so this seems to be some kind of "environmental" or other dative-like use. This is the only example of 3s > 1s =*nay*.
[18] This is one of two imperative sentences (both with =*ney*) in the dataset. It's not a normal imperative formation.
[19] JPH translated this sentence as "lo estoy aplastando" ('I am crushing it'). I'm assuming this was just scribal error.
[20] The form of the verb here (a bare stem) is unexpected, just as is the appearance of =*rey*. Maybe JPH just misheard.
[21] Maybe these are "environmental", e.g. 'The dog bit the man [we saw it]'.

arguments	*Series II*	*examples*	*Series I* ?
	=roy	Kokook=roy woshii' worooy-t-a. 'The dog bit the man' (CHM) bite.B=*roy* dog man-abs-acc	
2s intransitive	*='ay*	Xaay='ay 'oom chaaynok. 'You are not sick' not=*'ay* pron.2s be.sick.B	*='a='e* (usually realized as *='a='*)
	='ey	Hyaa='ey yaata-r. 'Now you're sleepy' now=*'ey* sleepy-abs	
3s intransitive	*=y*[22]	Paara=y menee' 'a-'aachen woshii'. 'That is this one's dog' that=*y* this 3sP-pet dog Wahii-ro=y. 'It's going to stink' stink-fut=*y* **Xaay=po=y pesaax**[23] **taame-t.** 'The sun is not yet up' (CHM) not=subj=*y* leave.B sun-abs	*='e*
3s > 3s transitive	*=y*	Hyaa=y na'uuk. 'He has now married her' (CHM) now=*y* marry.B	*='e*
3p intransitive	*=may*	Tavook=may shuuki namaake-nga. put.down.B=*may* ? middle-loc 'They put *shuuki* [shuukey 'tobacco'?] in the middle'[24]	*=me='e* (usually realized as *=m'e*)
	=moy	Chaarok=moy. 'They are dancing the patada dance' dance.patada.B=*moy* Wehee'-am=moy xaroo-ro. 'There will be two' two-pl=*moy* be-fut	
3p > 3s transitive	*=moy*	Ku'-kwii'ti'-am=moy mokaa-ro. 'The boys want to [will] kill it' red-boy-pl=*moy* kill-fut / **Ku'-kwii'ti'-am mokaa-ro=moy**.	*=ma' / =m='e*
3s > 2s transitive	*=mey*	**Pe'ee=mey worooy-t mokaa-ro.** 'This man will kill you' this=*mey* man-abs kill-fut / **Pe'ee worooy-t mokaa-ro=mey.**	*=m='e*
3s > 2p transitive	*=mey*	**Pe'ee worooy-t mokaa-ro=mey 'omooma'.** 'This man will kill this man-abs kill-fut=*mey* pron.2p you guys'	*=m='e*
3s > 2p environmental	*=may*	**Hamii-nga=m**[25]**=mey wakook?**[26] 'Where did it rain on you guys?' where-loc=?=*mey* rain.B	*=ha'=me* !!!
	=mey	'Ochoo'=mey. 'You are cold' (probably 'It is cold')[27] (CHM) cold=*mey*	*='a='* !!!
1s > 3s transitive	*=may*	Pwaana=may mokaanax. 'I killed him yesterday' (CHM) yesterday=*may* kill.B	*=n='e* !!!

[22] Undoubtedly I've missed examples of *=y* — it's hard to search for, since many other words end in *y*.
[23] This appears to be an accurate transcription of Merriam's "pe sah$\acute{\underline{ch}}$", but normally the simple stem *pesaa-* rather than the basic verb *pesaax* would be used in a sentence with the subjuntive clitic.
[24] JPH's translation of this is incomplete (only "they put").
[25] If indeed this is the mystery clitic *=m* (see fns. 8 and 11), in this example *=m* cooccurs with a Series II clitic and in fact comes before it. But maybe there are not really two *m*'s in this sentence.
[26] This is the only interrogative in the dataset.
[27] This seems most like an environmental inclusion of a second person argument.

arguments	*Series II*	*examples*	*Series I* ?
1s > 3p transitive	*=moy*	Huutok=moy paam. 'I see them.' (CHM) see.B=*moy* pron.3p	*=ne=me* !!!
2s > 1s transitive	*=neyay*[28]	Hyaa=neyay teroorkenax 'ooma'. 'Now you're breaking me'[29] now=*neyay* break.B pron.2s Huutok=neyay. 'You see me' (CHM) see.B=*neyay*	*=ne='a='*

3.2 *Analyzing the Series II clitics* As noted earlier, I suggested in Munro (2000) (following JPH, it seems) that it is probable that the final *y* of these clitics derives from the indicative mood clitic *='e.* But many questions remain.

• First, crucially, what determines the variation in the vowels in the various clitics? One hypothesis, considered in Munro (2000), is that these alternating vowels seem to reflect harmony with a preceding or following (stressed?) vowel, but there are many counterexamples to this idea in the data. However, the variation seems not to be random, since there are almost no cases of the same sentence recorded with more than one vowel choice. Here is one very rare such example:

(2) Too'avo-r=nay. / Too'avo-r=noy. 'I am blind' (JPH / CHM)
blind-abs=*nay* / blind-abs=*noy*

• Second, there are problems with the position of the Series II clitics. Some examples show a Series II clitic following a verb in non-second position:

(3) Ku'-kwii'ti'-am mokaa-ro=moy. 'The boys are going to kill it'
red-boy-pl kill-rut=*moy*

One hypothesis suggested for this (and the extremely rare comparable examples with Series I clitics) is that these items represent false starts of some sort, or maybe topicalization. But there are more examples like this with the Series II clitics. Do they suggest that the clitics are in a very preliminary stage of being reanalyzed as verb suffixes?

Two further examples suggest that Series II clitics may occur utterance-initially, e.g.

(4) Noy kuviinok. 'I am hungry'
noy be.hungry.B

As I suggested in Munro (2000), it seems very likely that JPH missed an initial adverb or other word in such sentences. But who knows?[30]

• Third, there are questions about the best analysis of the Series II clitics — does their final *-y* reflect the normal mood clitic *='e*? Perhaps the Series II clitics are at a preliminary stage of being reanalyzed as not including the normal mood (indicative) clitic *='e*. There are at least two imperative examples, two subjunctive ("wish") examples, and one interrogative ("question") example, (5):

(5) Hamii-nga=m=mey wakook? 'Where did it rain on them?'
where-loc=?=*mey* rain.B

[28] This is JPH's transcription, but CHM's might also be consistent with *=ney='ay.*
[29] JPH reports that " A man was once on top of a woman, & the woman said...".
[30] Jane Hill reminded me that there are many initial clitics in Serrano.

In addition, the 2s > 1s examples which illustrate a Series II combination *=ney=ay* suggest that the indicative clitic may occur twice, which is otherwise unprecedented. Thus, it seems that the Series II clitics that include *y* may be being reanalyzed as non-mood-indicating.

• Finally, as the Series II chart shows, there are a number of !!! uses of these clitics that are unparalleled for the well-behaved Series I set.

3.3 *Series II for the language learner* The Series II clitics present challenges for language teaching and learning.

• How does a speaker decide to use a Series II rather than Series I clitic? (There's some idea that the Series II clitics are "assertive", since it seems that they are used more often the Series I clitics to answer questions or respond to commands.)

• How does a speaker choose among the different vowel variants?

I have not yet tried to write a lesson about these clitics (though there are a number of examples of their use in our dictionary).

References

Bright, William. 1974. "Three Extinct American Indian Languages of Southern California." *American Philosophical Society Year Book*, pp. 573-74.

Buschmann, Johann Karl Eduard. 1865. *Die Sprachen Kizh und Netela von Neu-Californien dargestellt von Hrrn. Buschmann*. Berlin.

Hill, Jane H., and Kenneth C. Hill. 1968. "Stress in the Cupan (Uto-Aztecan) Languages." *IJAL* 34: 233-41.

Kroeber, A. L. 1907. *Shoshonean Dialects of California*. UCPAAE 4: 65-165.

McCawley, William. 1996. *The First Angelinos: The Gabrielino Indians of Los Angeles.* Banning, CA: Malki Museum Press-Ballena Press.

Merriam, C. Hart. [1903] Indian vocabularies. In McCawley (1996), pp. 237-64 (and partly consulted at the Bancroft Library, UC Berkeley).

Miller, Wick R. 1984. "The Classification of the Uto-Aztecan Languages." *IJAL* 50: 1-24.

Munro, Pamela. 2000. "The Gabrielino Enclitic System," in Eugene H. Casad and Thomas L. Willett, eds., *Uto-Aztecan: Structural, Temporal, and Geographic Perspectives.* Hermosillo, Sonora, Mexico: Universidad de Sonoral (Editorial Unison), pp. 183-201.

Munro, Pamela. 2002. "Takic Foundations of Nicoleño Vocabulary." In David R. Browne, Kathryn L. Mitchell, and Henry W. Chaney, eds., *Proceedings of the Fifth California Islands Symposium.* Santa Barbara, CA: Santa Barbara Museum of Natural History, pp. 659-668.

Munro, Pamela. In progress a. *Hyaare Shiraaw'ax 'Eyooshiraaw'a (Now We Are Speaking Our Language).* Ms.

Munro, Pamela, with input from the Gabrielino-Tongva Language Committee. In progress b. *Gabrielino/Tongva/Fernandeño Vocabulary with an English Index.* Ms.

Munro, Pamela, with John Johnson. 2001. "What Do We Know About Tataviam?" Ms.

Yo and the Rhetorical Interpretation

Takeshi Oguro
Chiba University of Commerce

1 Introduction

Rhetorical questions are generally considered to involve negative assertion unlike ordinary questions such as information-seeking questions. Since the seminal work of Sadock (1971), studies have been made to investigate the properties of these sentences, most notably by Han (2002).

Recently, Caponigro and Sprouse (2007) claimed that rhetorical questions are in fact ordinary questions in terms of syntax and semantics and it is pragmatics that makes the former different from the latter. To be exact, ordinary questions are rhetorical questions if both the speaker and the hearer know the answer and they are information-seeking questions if the speaker does not know the answer and he thinks that the hearer may. Caponigro and Sprouse's (2007) view is supported by studies on rhetorical questions in Japanese, represented by Fujii (2015), who examines the behavior of rhetorical questions ending with the expression *to iu no* and by Saito and Haraguchi (2012), who observe that the rhetorical interpretation is forced if the discourse particle *yo* is added to certain WH-questions.

In this paper, I would like to show, contrary to Caponigro and Sprouse (2007), that Japanese has a special type of question that is always rhetorical and, contra Saito and Haraguchi (2012), that the presence of *yo* in a question does not necessarily force the rhetorical interpretation.

2 Caponigro and Sprouse (2007): Rhetorical questions as ordinary questions

Caponigro and Sprouse (2007) claim that syntactically and semantically, there is nothing special about rhetorical questions and that they are distinguishable only in pragmatics. Their definitions of ordinary questions and rhetorical questions are given below.

(1) Definition of Ordinary Questions
An ordinary question is an interrogative clause whose answer is not known to the Speaker, but the Speaker thinks the Addressee may know it. An answer is required in order for the dialogue to be felicitous. Only the Addressee can answer.
(Caponigro and Sprouse 2007: 129)

(2) Definition of Rhetorical Questions
A rhetorical question is an interrogative clause whose answer is known to the Speaker and the Addressee, and they both also know that the other knows the answer as well. An answer is not required, but possible. Either the Speaker or the Addressee can answer.
(Caponigro and Sprouse 2007: 129)

This effect is shown in (3).

(3) QUESTION by the Speaker: Who cares about you?
ANSWER by the Speaker: Nobody.
by the Addressee: Nobody / Yeah, you're right.
(Caponigro and Sprouse 2007: 123)

The view of rhetorical questions as questions rather than as negative statements comes from (4).

Proceedings of *WECOL 2017*

(4) Negative statements never allow for an answer.
The Speaker: You should stop saying that Luca didn't like the party last night. After all, Luca was the only one that was still dancing at 3 am!
The Addressee or the Speaker: #Luca
(Caponigro and Sprouse 2007: 124)

With this background, let us review Saito and Haraguchi's (2012) analysis.

3 Saito and Haraguchi (2012) on *yo*

Saito and Haraguchi (2012) examine the roles played by sentence-final discourse particles in Japanese. They observe that WH-questions ending with the particle *yo* are forced to have the rhetorical interpretation.

(5) [$_{CP}$ Dare-ga soko-ni ik-u ka] yo
who-NOM there-to go-PRES Q] yo
Who will go there = No one will go there.'

One important point about *yo* is that it can be attached to various types of clauses.

(6) a. Taroo-wa soko-ni i-ta yo
Taroo-TOP there-at be-PAST yo
'Taroo was there.'
b. Kimi-wa soko-ni ik-e yo
you-TOP there-to go-IMP yo
'You go there.'
c. Soko-ni ik-oo yo
there-to go-EXH yo
'Let's go there.'

As shown in (6), *yo* can be attached to declaratives, imperatives, and exhortatives, which suggests that it does not select any particular clause type. Saito and Haraguchi assume that *yo* is an element in a Speech Act projection higher than CP and that it creates expressions with the speech act of assertion. Then, under this analysis, questions, when involving *yo* at the end of them, are interpreted as assertions. Since information-seeking questions cannot be asserted, they end up being interpreted rhetorically. Their analysis shares the insight of Caponigro and Sprouse (2007) according to which the rhetorical interpretation of questions is pragmatically determined.

4 The behavior of *ka yo* rhetorical questions

Saito and Haraguchi (2012) observe that WH-questions ending with *ka yo* are necessarily rhetorical, which they argue is due to the alleged assertive nature of the particle *yo*. This idea suggests that WH-*ka yo* questions are rhetorical only in pragmatics, which is in line with Caponigro and Sprouse's idea. It is therefore expected that they are syntactically and semantically ordinary questions. This section shows that, despite this expectation, *ka yo* questions behave quite differently from ordinary questions.

According to Caponigro and Sprouse (2007), rhetorical questions behave not as negative statements but as ordinary questions, with respect to answerability. Interestingly, *ka yo* rhetorical questions cannot be answered.

(7) a. Dare-ga soko-ni ik-u ka yo
who-NOM there-to go-PRES Q yo
'Who will go there = No one will go there.'
b. # Iya, daremo
no anyone
'No, no one.'

Adachi (2004), Sprouse (2007), and Fujii (2015) observe that WH-questions ending with *to iu no* are always rhetorical questions as in (8a). Importantly, they can be answered, as in (8b).

(8) a. Dare-ga soko-ni ik-u to iu-no?
who-NOM there-to go-PRES C saying-C
'After all, who will go there?' 'No one will go there.'
b. Iya, daremo
no anyone
'No, no one.'

Ordinary questions allow this sort of dialogue as well.

(9) a. Dare-ga soko-ni iki-mas-u ka?
who-NOM there-to go-POLITE-PRES Q
'Who will go there?'
b. Iya, daremo
no anyone
'No, no one.'

In this connection, it is interesting to see that *ka yo* rhetorical questions do not allow follow-up questions, while *to iu no* rhetorical questions and ordinary questions readily allow them.

(10) a. # Dare-ga soko-ni ik-u ka yo! John? Mary?
who-NOM there-to go-PRES Q yo John Mary
'No one will go there! John? Mary?'
b. Dare-ga soko-ni ik-u to iu-no? John? Mary?
who-NOM there-to go-PRES C saying-C John Mary
'After all, who will go there?' 'No one will come.'
c. Dare-ga soko-ni iki-mas-u ka? John? Mary?
who-NOM there-to go-POLITE-PRES Q John Mary
'Who will go there? John? Mary?'

There is another restriction on *ka yo* rhetorical questions. These are not compatible with 'which'-phrases, while *to iu no* rhetorical questions and ordinary questions have no problem in having them.

(11) a. * Dono hito-ga kur-u ka yo!
which person-NOM come-PRES Q yo
'Which person will come ? = No one will come.'
b. Dono hito-ga kur-u to iu-no?
which person-NOM come-PRES C saying-C
'After all, who the hell will come?' 'No one will come.'
c. Dono hito-ga ki-mas-u ka?
which person-NOM come-POLITE-PRES Q
'Which person will come?'

Another point which distinguishes *ka yo* rhetorical questions is that they license a strict type of negative polarity items such as *daremo* 'anyone.' As shown in (12), *daremo* is only allowed in negative contexts (McGloin 1976).

(12) a. Daremo ko-na-i
anyone come-NEG-PRES
'No one will come.'

b. * Daremo kur-u
anyone come-PRES
'No one will come.'

c. * [Daremo kure-ba] boku-wa uresii
anyone come-if I-TOP happy
'If anyone comes, I am happy.'

d. * Daremo ki-ta to-wa odoroki da
anyone come-PAST C-TOP surprise COP
'It is surprising that anyone came (at all).'

The contrast in (13) indicates that *daremo* is licensed in *ka yo* rhetorical questions but not in *to iu no* rhetorical questions or in ordinary questions.

(13) a. Daremo kur-u ka yo!
anyone come-PRES Q yo
'No one will come!'

b. * Daremo kur-u to iu-no?
anyone come-PRES C saying-C
'After all, anyone will come?' 'No one will come.'

c. * Daremo ki-mas-u ka?
anyone come-POLITE-PRES Q
'Will anyone come?'

Another point where *ka yo* rhetorical questions differ from *to iu no* rhetorical questions and ordinary questions is that they do not tolerate the mention-some interpretation, which is indicated by the addition of *tatoeba* 'for example,' while the latter two types of questions are fine with it. This contrast is illustrated in the following.

(14) a. * Tatoeba dare-ga kur-u ka yo!
for.example who-NOM come-PRES Q yo
'No one, for example, will come!'

b. Tatoeba dare-ga kur-u to iu-no?
for.example who-NOM come-PRES C saying-C
'After all, who, for example, will come?' 'No one will come.'

c. Tatoeba dare-ga ki-mas-u ka?
for.example who-NOM come-PRES Q
'Who, for example, will come?'

Ka yo rhetorical questions are further subject to some kind of tense restriction in the sense that they fail to involve past tense. *To iu no* rhetorical questions, on the other hand, are exempt from it.

(15) a. * Dare-ga kinoo ki-ta ka yo!
who-NOM yesterday come-PAST Q yo
'Who came yesterday? = No one came yesterday.'

b. Dare-ga kinoo ki-ta to iu-no?
who-NOM yesterday come-PAST C saying-C
'After all, who the hell came yesterday?' 'No one came yesterday.'

c. Dare-ga kinoo ki-masi-ta ka?
who-NOM yesterday come-POLITE-PAST Q
'Who came yesterday?'

As shown above, *ka yo* rhetorical questions should be treated differently from ordinary questions, while *to iu no* rhetorical questions should be regarded as behaving on a par with ordinary questions.

5 The behavior of *mono ka* rhetorical questions

Saito and Haraguchi (2012) observe that WH-questions ending with *ka yo* are necessarily rhetorical, which they argue is due to the alleged assertive nature of the particle *yo*. This idea suggests that WH-*ka yo* questions are rhetorical only in pragmatics, which is in line with Caponigro and Sprouse's idea. It is therefore expected that they are syntactically and semantically ordinary questions. This section shows that, despite this expectation, *ka yo* questions behave quite differently from ordinary questions.

It is very important to notice that the non-ordinary behavior of *ka yo* rhetorical questions mirrors that of *mono ka* rhetorical questions, which is examined in Oguro (2014, 2015, and 2017). As shown in (16), all the properties of *ka yo* rhetorical questions seen in the previous section are shared by *mono ka* rhetorical questions.

(16) a. Dare-ga kur-u mono ka!
who-NOM come-PRES C Q
'No one will come!' (RQ)

b. * Dare-ga kur-u mono ka! John? Mary?
who-NOM come-PRES C Q John Mary
'No one will come! John? Mary?'

c. * Dono hito-ga kur-u mono ka!
which person-NOM come-PRES C Q
'No one will come!'

d. Daremo kur-u mono ka!
anyone come-PRES C Q
'No one will come!'

e. * Tatoeba dare-ga kur-u mono ka!
for.example who-NOM come-PRES C Q
'No one, for example, will come!'

f. * Dare-ga kinoo ki-ta mono ka!
who-NOM yesterday come-PAST C Q
'No one came yesterday!'

It is proposed in Oguro (2017) that *mono ka* rhetorical questions have the following structure.

(17) $[_{CP}\ OP_{[NEG]}\ldots\ [_{FINP}\ TP_{[subjunctive]}\ [_{FIN}{}^{0}$ mono$]]\ [_{C}{}^{0}{}_{[RQ]}$ ka$]]$

I assume with Rizzi (1997) and many others that the CP-domain may involve various functional projections. In (17) the domain contains two projections, the higher one headed by *ka* and the lower one headed by *mono*.

The C-head *ka* in (17) is assumed to contain the RQ-feature, which is present in *mono ka* rhetorical questions, but not present in ordinary questions or in *to iu no* rhetorical questions, both of which are assumed to contain the OQ-feature. The RQ-feature is responsible for the unanswerability of such sentences.[1] The spec of the RQ head is occupied by the negative operator, which is responsible for the negative interpretation and the licensing of strictly negative polarity items such as *daremo* 'anyone'. The WH-phrase in such a question is interpreted as a negative quantifier rather than as an existential quantifier, so the mention-some interpretation is not possible.

As for the impossibility of 'which'-phrases in these rhetorical questions, I assume that the negative operator, which guarantees the WH-phrase to be interpreted as a negative quantifier, is strong enough to deny the presence of a contextually defined set of possible answers for the WH-phrase. Thus negation applies not just to a specific group of individuals but it applies to all individuals, which makes any specific group of individuals irrelevant. Since 'which'-phrases require the presence of such a set, a conflict arises in (16c).

[1] Perhaps the term RQ might be confusing, since *mono ka* rhetorical questions are not typical questions but should be treated as negative assertions.

Mono is assumed to head the lower functional projection, that is, the Finite projection. It selects TP. In most cases, it is headed by *no*, as assumed by Saito and Haraguchi. In *mono ka* rhetorical questions, it is headed by *mono*. I assume that the Finite head *mono* in *mono ka* rhetorical questions selects subjunctive TP, which is incompatible with specific episodic tense, which captures the deviance in (16f). Thus, (16a) means that no one in any imaginable world would come.[2]

Given this set of facts, it seems fair to say that *ka yo* rhetorical questions are a subclass of *mono ka* rhetorical questions, with the Finite head *mono* being silent. The idea of silent *mono* is supported by the following.

(18) Dare-ga ik-u ka!
who-NOM go-PRES Q
'No one will go!' (RQ)

This example is only possible as a rhetorical question.[3] Thus, *ka yo* rhetorical questions are rhetorical even without the discourse particle *yo*, which makes dubious the claim by Saito and Haraguchi (2012) that the particle *yo* forces the rhetorical interpretation of questions.[4]

6 Non-rhetorical questions ending with *yo*

Despite Saito and Haraguchi's claim, there are instances of questions ending with the particle *yo* which are not necessarily rhetorical. As observed in Nitta (1991), non-polite WH-questions ending with *ka* such as (18) are rhetorical, but those ending with the copula *da* are ordinary questions. As illustrated in (19), this type of questions can involve *yo*, retaining the interpretation as ordinary questions (Davis 2011).

(19) a. Dare-ga soko-ni ik-u n(o) da (yo)?
who-NOM there-to go-PRES FIN COP yo
'Who will go there?'

b. Dare-ga soko-ni ik-u n(o) da (yo)? John? Mary?
who-NOM there-to go-PRES FIN COP yo John Mary
'Who will go there? John? Mary?'

c. Dono hito-ga soko-ni ik-u n(o) da (yo)?
which person-NOM there-to go-PRES FIN COP yo
'Which person will go there?'

d. * Dare-ga dokoemo ik-u n(o) da (yo)?
who-NOM anywhere go-PRES FIN COP yo
'Which person will go anywhere?'

e. Dare-ga kinoo soko-ni it-ta n(o) da (yo)?
who-NOM yesterday there-to go-PAST FIN COP yo
'Who went there yesterday?'

[2] These properties of *mono ka* rhetorical questions are strongly reminiscent of the behavior of WH-*the hell* questions, which are observed by Pesetsky (1987) to be incompatible with *which*-phrases, and Den Dikken and Giannakidou (2001) to be forced to have the rhetorical interpretation when they involve a modal expression like *would*, a sign of subjunctive tense. They assume that WH-the expressions are licensed a nonveridical operator in the C-domain and that the combination of this operator and the subjunctive tense forces the rhetorical interpretation. In *mono ka* rhetorical questions, on the other hand, it is the negative operator that forces the negative interpretation and the subjunctive tense.

[3] This observation goes back to Nitta (1991), who observes that WH-questions without a politeness marker are not acceptable as information-seeking questions but fine as rhetorical questions.

[4] *Mono ka* rhetorical questions are compatible with *yo*.

(i) Dare-ga kur-u mono ka yo!
who-NOM come-PRES C Q yo
'No one will come!' (RQ)

These examples contain the Finite complementizer *no*, which selects ordinary TP, and they behave as ordinary questions. Here, the presence of the discourse particle *yo* does not affect their interpretation, contrary to Saito and Haraguchi's analysis.

There is another kind of question which ends with *yo* but keeps the interpretation of ordinary questions. Unlike the WH-counterparts, non-polite *yes-no* questions ending with *ka* are allowed as ordinary questions and the addition of the particle *yo* does not affect the status of these examples as ordinary questions. This is illustrated in the following paradigm.

(20) a. Taroo-wa ku-ru no ka (yo)?
Taroo-TOP come-PRES FIN Q yo
'Will Taroo come?'

b. * Daremo ku-ru no ka (yo)?
anyone come-PRES FIN Q yo
'Will anyone come?'

c. Taroo-wa ki-ta no ka (yo)?
Taroo-TOP come-PAST FIN Q yo
'Did Taroo come?'

On par with the examples in (19), the ones in (20) also involve the Finite head *no*, rather than *mono*, and their behavior patterns with that of ordinary questions. What is important is that their status as ordinary questions remains intact even with the presence of the discourse particle *yo*.

What we have seen in this section and the previous section is that the presence of *yo* does not affect the interpretation of questions. The observation in this section in particular shows that information-seeking questions can involve *yo*, which refutes Saito and Haraguchi's analysis of the particle *yo* as creating expressions with the speech act of assertion.[5]

7 The function of *yo*

This section aims to clarify the role played by the particle *yo*. Saito and Haraguchi claim that it has to do with assertion, but that is not an accurate characterization, which was shown in the previous sections. As provided in the following, there is more evidence for the irrelevance of assertion in the distribution of *yo*, which will be shown to elucidate its function.

As exemplified in (21), there are various uses of *yo*, which have nothing to do with assertion.

(21) a. ano yo b. hora yo c. titi yo, d. Yo!
well yo look yo father yo yo
'well' 'look' 'father' 'Hi!'

Yo can be attached to a filler as in (21a), to an interjection as in (21b), to a vocative phrase as in (21c), and it can stand alone and be used as a greeting as in (21d).[6] In all these cases, the particle is used by the speaker to attract the hearer's attention. These examples tell us that it is simply a sign of the speaker addressing the hearer, roughly meaning "I am talking to you."

The suggested view of *yo* is consonant with the speaker's impatience which can be detected in questions ending with this particle. *Yo*-questions are often uttered when the hearer is hesitant to respond or keeps giving evasive answers. Questions are generally uttered to serve two discourse-related purposes: one is to address the hearer and the other is to solicit a response from him. The hearer has two jobs: one is to receive the question and the other is to respond to the question by providing an answer. The role of *yo* is

[5] Following Oguro (2015), I assume that *mono ka* rhetorical questions always involve Speech Act Phrase, hence unembeddable. I agree with Saito and Haraguchi that *yo* belongs in that domain, but I claim that it does not create any type of speech act.

[6] These expressions are assumed to be contained in a Speech Act Phrase. Note that a speech act head selects an utterance, which does not necessarily have to be a proposition or a question but can be a greeting or an act of addressing someone.

only to draw the attention of the hearer and it does not include inviting an answer. Thus, ordinary questions with *yo* sound like they are uttered to the hearer just to catch the hearer's attention, without the intention of soliciting an answer. [7,8] This sort of question make sense if the speaker wants to make sure that the hearer is listening. Thus, when the speaker utters this type of question in a normal context, where the hearer is clearly listening, he ends up implying that the hearer is not listening to him, which makes the question sound rude.

There is another case which helps understand the function of *yo*.

(22) Dono hito-ga ki-ta no ka?
which person-NOM come-PAST FIN Q
'Which person came?'

This is a non-polite WH-question, and as noted by Nitta (1991), it does not have the information-seeking reading. Despite Nitta (1991), however, it also lacks the kind of rhetorical interpretation that is available in *mono ka* rhetorical questions, as shown by the presence of the 'which'-phrase, the past form of the verb, and the Finite head being *no*, not *mono*. This example, however, is acceptable as a self-addressed question, where the speaker asks himself but not the hearer. Interestingly, the addition of *yo* makes the example degraded, even under the self-addressed question reading.

(23) ?* Dono hito-ga ki-ta no ka yo
which person-NOM come-PAST FIN Q yo
'Which person came?'

This sentence is deviant. It is intended as a self-addressed question and at the same time, the speaker utters it to draw the hearer's attention without attempting to solicit information. Thus the pragmatic property of *yo* does not go with the intention of this question, leading to deviance. The speaker tries to catch the hearer's attention by asking a question, but the question is not directed to him.[9] This effect is not predicted by Saito and Haraguchi's approach. It would incorrectly predict (23) to be rhetorical due to the presence of this particle. Thus, their view of this particle, as theoretically intriguing as it seems, is empirically problematic.

8 Conclusion

In this paper, I examined Saito and Haraguchi's view of the relation between the rhetorical interpretation of certain WH-questions and the discourse particle *yo*, as well as the idea that the rhetorical interpretation of questions is pragmatically determined. It was shown that there are cases of non-rhetorical questions ending with *yo* and it was suggested that the *ka yo* rhetorical questions dealt with by Saito and Haraguchi are in fact a subcase of what I call *mono ka* rhetorical questions, which are always rhetorical even without the particle. I also suggested that the role of *yo* is merely to catch the attention of the hearer. This view captures the awkwardness found in self-addressed questions ending with *yo*.

9 References

Adachi, Taro. 2004. Rhetorical questions in Japanese. *Memoirs of Kyoto Tachibana Women's University*: 31, 35-50.
Caponigro, Ivan, and Jon Sprouse. 2007. Rhetorical questions as questions. In Puig-Waldmüller, Estella (ed.), *Proceedings of Sinn und Bedeutung 11*. Barcelona: Universitat Pompeu Fabra, 121-133.

[7] A more accurate translation of (19a) would thus be "I said, who will come!"
[8] Another kind of questions which the speaker utters only to draw the hearer's attention but not to invite an answer are greetings such as *How do you do?* or *What's up?*, to which the hearer usually reply with the same questions.
[9] Japanese has a special form of questions ending with *yara*, in which the speaker asks himself but not the hearer (Oguro 2016). They also fail to be accompanied yo, as illustrated in (i).

(i) Dare-ga kur-u no yara (?*yo)
who-NOM come-PRES FIN yara yo
'I wonder/don't know who will come.'

Davis, Christopher M. 2011. *Constraining interpretation: sentence final particles in Japanese*. Doctoral dissertation. University of Massachusetts.

Dikken, Marcel den and Anastasia Giannakidou. 2002. From *hell* to polarity: "Aggressively non-D-linked" WH-phrases as polarity items. *Linguistic Inquiry* 33: 31-61.

Fujii, Tomohiro. 2015. Wh-Shuji gimon ni okeru shima no koka nitsuite no oboegaki [A Note on the Island Effects in Wh-Rhetorical Questions], in *The deep/true phase of language essays presented to Professor Yukio Oba on the occasion of his retirement from Osaka University*, 267-273. Tokyo: Eihosha.

Han, Chung-hye. 2002. Interpreting interrogatives as rhetorical questions. *Lingua* 112: 201-229.

McGloin, Naomi Hanaoka. 1976. Negation. In *Syntax and semantics 5: Japanese generative grammar*. Shibatani, Masayoshi (ed.), 371-419. New York: Academic Press.

Nitta, Yoshio. 1991. *Nihongo no modariiti to ninsho*. Tokyo: Kenkyusha.

Oguro, Takeshi. 2014. Negation in certain rhetorical questions in Japanese. *Proceedings of the Florida Linguistics Yearly Meeting (FLYM) 1*.

Oguro, Takeshi. 2015. WH-questions in Japanese and speech act phrase. *Linguistica Alantica*. 34 (2): 89-101.

Oguro, Takeshi. 2016. Speech act phrase and conjectural questions in Japanese. *Proceedings of the Eighth Formal Approaches to Japanese Linguistics Conference (FAJL8)*, 97-108.

Ougro, Takeshi. 2017. Properties of *mono ka* rhetorical questions. paper read at the 13th Workshop on Altaic Formal Linguistics 13 (WAFL 13).

Pesetsky, David. 1987. WH-in-situ: movement and unselective binding. In *The representation of (in)definiteness*. Reuland, Eric and Alice ter Meulen (eds.), 98-129. Cambridge, Ma: MIT Press.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of grammar: handbook of generative syntax*. Liliane Haegeman, (ed.), 281-337. Dordrecht: Kluwer.

Sadock, Jerold. 1971. "Queclaratives," *Papers from the Seventh Regional Meeting of the Chicago Linguistic Society*. pp. 223-232.

Saito, Mamoru and Tomoko Haraguchi. 2012. Deriving the cartography of the Japanese right periphery: the case of sentence-final discourse particles. *Iberia* 4: 104-123.

Sprouse, Jon. 2007. Rhetorical questions and wh-movement. *Linguistics Inquiry* 38: 572-580.

The Language of Ritual: Selected Tongva Etymologies

Brent Scarcliff
Gabrielino-Tongva Language Committee

1 Introduction

Drawing upon the painstaking work of Dr. Pamela Munro, I am current preparing an etymological dictionary of the Gabrielino-Tongva language (Scarcliff 2017, in progress), including a preliminary reconstruction of Proto-Takic. In an attempt to better understand the many ritual-oriented comparative data sets where the phonological correspondences seem clear, but the semantic relationships do not, I have devised a simplified model of the Takic ritual system, summarized so:

OPPOSITION half	RESOLUTION whole	OPPOSITION half
THE DARK [1] *tu-ku- 'black'	THE NEUTRAL [2] *ti-pa- 'white'	THE BRIGHT [3] *pa-ha/i- 'dawn; shine; bright'
THE EARTH [4] *ti- 'earth'	THE CREATOR [5] *su- 'star; spirit'	THE SKY [6] *pa-ha-ve-t 'eastern power'
THE EARTH PEOPLE [7] *tu-ku- 'moiety marker'	THE STAR PEOPLE [8] *ʔisa-ta- 'White Coyote'	THE SKY PEOPLE [9] *pa(-ha/i-) 'moiety marker'
THE DARK ONE [10] *tu-ku-ʔu-t 'Wildcat; mascot'	THE CONDUIT [11] *ti-pa-ʔa-t 'Pinyon; world tree'	THE BRIGHT ONE [12] *paha-ʔa-t 'Bighorn; mascot'
THE FIRST WOMAN [13] *tu-ku- 'woman'	THE SHAMAN [14] *pu- 'blow'	THE FIRST MAN [15] *wi(ʔ)-wi(ʔ)-wɨ-t 'brainy one'

By my definition, a moiety is one half of any culture which ritually divides itself into two. Moiety-based societies are common the world over, with famous examples from Africa, Australia, Southeast Asia, and, of course, the Americas:

> "We know that a number of Eastern Woodlands First Nations […] divided communities into moieties (halves) called Sky and Earth, or clans such as Bear, Wolf, and Turtle […]" (Kehoe 2002)

> "[…] the Central Sierra Miwok are divided into exogamic moieties with paternal descent, usually spoken of as kikua (water side) and tunuka (land, or dry, side). Frequently the former are referred to as 'bullfrog people' (lotasuna) and the latter as 'bluejay people' (kosituna)." (Gifford 1916)

It is clear that the ancestral Takic peoples likewise divided their world into two complementary but also conflicting halves.

> "The Serrano are organized on the basis of exogamous totemic moieties. One moiety is called wahilyam, and has as totems coyote (wahil), the chief totem for which the moiety is named, buzzard (widukut), and wolf (wanats). The other moiety is called tukum, and has as totems wild cat (tukum), the chief totem for which the moiety is named, crow (gatcawa), and mountain lion (tukutcu)." (Mason 1918)
> "The moon was the only woman among all Mūkat's creatures. Every morning she would go away from the ceremonial house to a clean sandy place, where with woven grass string she showed all the creatures how to make cat's cradles. Then she would put one group of people on one side and

> say, "you are coyote people," and the others she would call wildcat people. She told the coyote people to sing against the wildcat people as though they were singing enemy songs. Then the wildcat people would begin to dance; then they would do it the other way around." (Agua Caliente Cultural Museum 2013)

> "Every lineage belonged to one of the two moieties, which were named either 'wildcat' or 'coyote'." (McCawley 1996)

As these observations suggest, moieties are the key to understanding many kinship systems, serving as the institution of exogamous marriage's reason for being. Compare, for example, Tongva ʔa-shó:ʔ 'other' and ʔa-shó:ʔ-n 'husband; wife' to Proto-Malayo-Polynesian *hipaR 'sister-in-law; the other side of the river' (Blust & Trussel 2017). Both terms reflect the custom of marrying outside your group. Moieties are, however, equally relevant to a culture's mythology, where they serve to explain the natural dichotomies we find in nature. Consider the Apinayé people of Brazil (da Matta 1982), who similarly divide their world into two halves: (Moon, Night, Women, Water, Winter, Left, West, Black, Wet Season, Fierce, Rhea, Wrong) vs. (Sun, Day, Men, Fire, Summer, Right, East, Red, Dry Season, Tame, Tapir, Right). Or the Han people of China (Wang 2012), who do likewise: (Yin, Dark, Moon, Negative, North, Female) vs. (Yang, Bright, Sun, Positive, South, Male).

The basic model above, which is subject to extensive revision, is based upon Richard Applegate's 1978 structuralist analysis of Luiseno mythology and my own research into Takic ritual and story and. One obvious question springs to mind. If a moiety is, by definition, one of two, why are there three columns? I have two answers. First, Takic ritual itself is tripartite in nature. There are three sacred colors. Three rings in a sacred sand painting. Three social classes. And so on. Second, according to the Lévi-Strauss structuralist theory of mythology, the essential purpose of ritual (which may be thought of third, mediating force) is to reconcile the two opposing forces. If we consider Tongva's three ritual colors (red, white, and black), for example, we may be reminded of the symbolic tripartite color scheme of Indo-European mythology: the red of the sky/warrior function, the white of the priestly/kingship function, and the blue-green of the earth/fertility function (see Puhvel 1987).

Below I suggest several new etymologies based upon this model. Unless otherwise noted, all Proto-Takic reconstructions are my own and should be considered tentative. For convenience, I let the Kitanemuk absolutive endings stand in for their corresponding proto-forms, in keeping with my current view that it may be the most conservative Takic language in this regard. Unless otherwise noted, all other language data come from the standard reference works, namely: Seiler & Hioki 2006 for Cahuilla, Hill & Nolasquez 1973 for Cupeño, Anderton 1988 for Kitanemuk, Elliott 1999 for Luiseño, SLRP 2012 for Serrano, Munro 2017 for Tongva, and Stubbs 2011 for Proto-Uto-Aztecan. For clarity's sake, I have slightly adjusted the original orthographies in these works and added my own segmentation.

2 The Dark

Takic ritual requires the use of black, white, and red, perhaps better labeled dark (e.g. black, blue, and green), neutral (e.g. white, grey, and brown), and bright (e.g. red, yellow, and orange), respectively.

> "The painting was made with red and yellow paint, paesul and navyot, ashes for white, and charcoal for black, on the ground which formed the background of the painting. The entire picture, which was circular and represented the world, was called torokhoish." (DuBois 1908)

> "Female mourners cut their hair just below the ears and put black tar on their foreheads, cheeks, and chin." (McCawley 1996)

Darkness is the chief attribute of Sister Earth, whose other qualities include being Below, Dead, Nocturnal, Oceanic, and Western:

Proto-Takic *tu-kuC-t ~ *tu-huC-t 'black clay; black paint; the color of femaleness'
Tongva tú:xo-t 'black clay'
Tongva tokú:-r-aw-t 'pretty'

Kitanemuk tuhu-t ~ tuhu-c ‘black paint’
Cf. PUA *tu-L ‘charcoal’ (entry 241)
Cf. PUA *tuki ~ *tuka ~ *tuku ~ *tuhu ‘dead fire; dark; black; night’ (entry 240)

3 The Neutral

In contrast to Sister Earth’s absence of color (dark black) and Brother Sky’s surfeit of color (bright red), the First Creator’s colors (white, brown, and grey) are decidedly neutral:

> “In the Luiseño creation myth, xwaykit piwkit 'whitish, greyish’ immediately precedes the creation of Earth and Sky.” (Applegate 1979)

In a practice known as geophagy, pregnant women may consume kaolin, also known as chalk, white dirt, or white clay (see Hunter 1973). In Takic terms, white clay is mankind’s first food.

> “Then it was told them that in all time to come they must have fiestas for the dead as they had done for Ouiot. And they must begin to kill and eat for food. Until this time they had never eaten flesh or grains, but had lived on [white] clay.” (DuBois 1908)

Proto-Takic *tɨva-č ~ *tɨvi-c ‘white; white clay; the color of creation’
Cahuilla tévi-š-nek ‘white’
Luiseño tó:vi-š ‘white clay’
Luiseño tó:va-l ‘white clay’
Luiseño -tó:vi ‘semen; the substance that squirts out of a rattlesnake when stepped on’
Luiseño tóv-ka/i ‘to search for white clay’
Luiseño tó:vu-k-il ‘month name; white clay; first food of the northerners’
Luiseño tóv-la/i ‘to bear a child’
Luiseño -tó-vli ‘frog spawn; fish spawn’
Tongva tó:ve-y ‘white clay’
Tongva tová:-r ‘earth’
Tongva tová:-ŋa-r ‘world, year, church’
Kitanemuk tɨvi-č ‘lime, for painting white stripes on grave poles’
Kitanemuk tɨva-č ‘land; earth; ground; year; world’

4 The Bright

Harrington recorded an unusually rich set of Tongva ‘dawn’ words, making it possible to reconstruct the underlying etymon in more detail:

Proto-Takic *pá:ha- ‘to dawn; to shine’ ~ *pahí:-c ‘dawn’ ~ *pá: ‘to shine (reduced form)’
Proto-Takic *pa-ʔapka-č ‘sunflower; dawn-sunflower’
Cahuilla -pá:y- ‘to dawn; to sit up all night’
Cahuilla paha-hūwit ~ su:-wi-t ‘a star’
Cahuilla pa:yi-š ‘dawn; morning’
Tongva pá:ha-t-kom-ok ‘to shine’
Tongva pahí:-y-n-ok ‘to dawn’
Tongva ʔa-pá:ha-t-kom-e-n ‘its light’
Tongva ʔa-pá:hi-y-n ‘its day’
Tongva pahí:-y-o-t ‘the morning star’
Tongva pá:-xa-r ‘wild sun-flower’
Kitanemuk pahi-kyɨ-t ‘the morning star; the rainbow goddess; a certain color’
Serrano ku-pa:ʔ-n ‘to shine (like the stars)’
Cf. SUA *pacay ‘shine’ (entry 2233)
Cf. NUA *pahay ‘dawn’ (entry 2234)
Cf. PUA ʔapkaC ‘sunflower’ (entry 2239)
Cf. Proto-Numic *ta-pa(i) ‘(day)light, sun’

Cf. Proto-Takic *kwa-č ‘red magic’

5 The Earth

As we have seen, the qualities associated with Sister Earth include being Down, Under Ground, and Under Water.

> “My heart is lost, lost. My heart sets, sets. My heart goes to the other world. My heart goes to the other world. My heart goes to the ocean foam. My heart goes to the ocean foam.” (Hill & Nolasquez 1973)

Proto-Takic *ti-č ‘deep; underground; underwater; the location of the underworld’
Cahuilla té-l-me-ki-š ‘the place where dead people live’
Cupeño té-l ‘comet’
Luiseño tó-l-mu-l ‘hell; underground’
Luiseño tó:-ŋa-x ‘down; downhill’
Luiseño pá:-ti-ŋa-l ‘dolphin; porpoise’
Tongva tó:-r-kwa-ve-t ‘the middle of the earth’
Tongva tó:-ŋkoʔ ‘down’
Tongva to-ró:-ve-m ‘feather headdress wearers who take care of the world’ [reduplicated]
Tongva tu-tú:-ŋʔa-ve-t-am ‘ones from below’ [reduplicated]
Kitanemuk tɨ-ʔ ‘roast’
Kitanemuk tɨh-pok-t ‘mole’
Kitanemuk tɨʔ-ŋa-ʔ ‘deep, said of water or a canyon’
Cf. Tübatulabal tɨ-ŋgi:-l ‘rock ledge’
Cf. PUA *tɨ- ‘below’ (entry 700)
Cf. Proto-Takic *tukuh-pa-ča ‘night sky; night-shine; heaven; bead; iron meteor; metal’

6 The Creator

According to Takic tradition, the Creator Ab Initio is the White Milky Way, from whence the People come and to which they may expect to return:

> “The stars were those of the First People who went up into the sky in the attempt to escape from death […]. (DuBois 1908)

> “A treble circle was drawn on the ground, with an opening on the side facing the north. The outer circle represented the Milky Way, the middle one night, and the inner one blood. A small pit, representing "hell," was made in the center of the treble circle, and figures of animals were made between this hole and the inner circle, as indicated in the illustration.” (Sparkman 1908)

> “One of the songs of Kwinamish, already given, begins: Tomamik Yula Wanawut poponakala ponyarakala auma, to the north the spirit in carefully woven strings remains tied. The striking peculiarity in the Luiseño use of sacred terms, the doubling of the word, has been referred to. Wanal Wanawut has been explained. Yula Wanawut has almost exactly the same significance. Yula means spirit, and literally head or hair. It is possible that in ancient times the object Wanawut was made of hair, as were the bracelets and anklets used in the girls’ ceremony.” (DuBois 1908)

The tests required for the newly deceased to reach heaven include fetching a white feather (symbolic of the flight to heaven) from atop the gravepole, splitting a hair (symbolic of the way to heaven) from end to end, drawing a star map (a map of heaven), and slaying a herd of deer disguised as beetles (both Earth Side entities). (See McCawley 1998)

Takic ritual suggests our hearts are stars, strung like beads across the night sky. Our bodies are seeds and shells, housing our spirits. And our hair is a basket, web, or net (e.g. Luiseño wá:na-l ‘the net connecting heaven and earth’), binding creation together.

Proto-Takic *su-‘star; heart; seed; shell; hair; spirit’
Luiseño ʂúʔ-la ‘star’
Luiseño ʂú:-n ‘heart; core; soul; internal organ’
Tongva šú:-r ‘star’
Tongva -šú:-n ‘heart’
Tongva -šú:-nox ‘single hair’
Tongva -šú:-ha-r ‘hairbrush’
Kitanemuk huʔ-č ‘star; landsnail’
Kitanemuk hu-na-c ‘heart; spirit; middle’
Cf. PUA *su-na ‘heart; inner part; seed’ (entry 1165)
Cf. PUA *su-’u ~ *su-wa ‘star’ (entry 2169)

In the Takic world, deer and rabbits (because of the sustenance their deaths provide) and birds and insects (because of their ability to fly) form a kind of messenger class, acting as conduits between this world and the next. This same root appears to serve as the base for many of their names.

Proto-Takic *su-kah-t ‘spirit-being; star-being; a being that moves between worlds’
Tongva šú:-ša-r ‘deer’s hoof rattle’
Tongva šu-ká:-t ‘deer’
Tongva šo-pé:reʔ ‘dragonfly'
Tongva šo-ká:-wo-t ‘flying squirrel’
Tongva šo-ká:ʔ ‘wasp’
Tongva šu-ʔí:-t ‘jackrabbit’
Kitanemuk hukah-t ‘deer; water spider’

7 The Sky

The Takic religious lexicon suggests the first light of day has a dangerous power, making use of the same root (*pá:ha) identified in Section 3 above:

Proto-Takic *pá:ha-viC-t ‘of the dawn; ritual power’
Luiseño pá:ha-vi-t ‘a great thing what made this world; glutton’
Luiseño ʔawáy pá:ha-vi-t ‘said of someone who does something excellent’
Kitanemuk paha-vi-t ‘dream helper; bear; rattlesnake; ghost; poison’
Serrano pa:hä-vi-t ‘doctor’s thing; holy one; powerful one; sacred power’
Tongva pá:ha-ve-t ‘morning star’
Tongva Pabavit ‘The First Man’ (Kroeber 1925)
Cf. SUA *pahatu ‘poison’ (entry 1696)
Cf. Nahuatl pahii ‘to take medicine or poison’

8 The Earth People

In the Takic world, every one of the People (i.e. every animal, plant, and natural phenomenon) belongs to either the Earth Side or the Sky Side.

> “The animate world was divided into two classes, hence an animal or plant name must belong to one or the other.” (Strong 1929)

Many of the names of the Earth People evidently share that same root (*tuku-) discussed in Section 1 above:

Proto-Takic *tuku- ‘Earth moiety marker’
Cahuilla túku-t ‘wild-cat’
Cahuilla túk-we-t ‘mountain lion’
Cupeño tuku-t ‘wildcat’

Cupeño tukú-pu-we-t ‘woodpecker’
Luiseño tuka-púy-puy ‘ant lion’
Luiseño túk-va-l ‘sea otter’
Luiseño túku ‘song word of unknown meaning, said to be in frog language’
Luiseño tukurú:ka/i ‘be dotted; make dots’
Luiseño túku:-pa-wu-t ‘oriole; kingbird; butcher bird’
Luiseño túk-yapa-l ‘screech owl’
Tongva tokú:-wa-r ‘raccoon’
Tongva tukú:-t ‘wildcat’
Tongva tú:ko-ka-r ‘whirlwind’
Fernandeño tukú:-r ‘panther’ (Kroeber 1925)
Kitanemuk tuku-ahpa-č ‘mockingbird’
Kitanemuk tuku-ku ‘tarantula; a pahavit’
Kitanemuk tuku-vɨ-č ‘sea otter’
Cf. Tübatulabal tu:ga-iya:-l ‘the actual black bird’ (Voegelin 1958)
Cf. Tübatulabal tu:ga-ya-ya:-l ‘the mythical black bird’ (Voegelin 1958)
Cf. PUA *tuku ~ *tuka ~ *tuki (entry 240)

As an example of an unrelated language family with a similar approach, Proto-Austronesian likewise employs a prefix (*qali- ~ *kali-) to mark certain animals having supernatural qualities (see Blust & Trussel 2017).

9 The Star People

As the titular head of the Sky Moiety in most Takic traditions, we would expect Coyote to represent the Bright World. In the Piman tradition, however, he evidently represents the White World, as I categorize him here:

> “The Pima world as a whole is divided into the duality of Coyote, white, cold and Buzzard, red, hot.” (Shimkin 1941)

Like Anansi, Loki, Maui, and Raven, Coyote is a transfunctional figure: sub-creator, clown, mankind-helper, evil-doer, and trickster. He is the fly in the ointment. The yeast in the bread. The wrench in the works. As such, he has several ‘avatars’, one for the Dark Earth Side, one for the Bright Sky Side, and one for the Neutral Creator Side.

> “He was called paya isil ‘water coyote’, tamiya isil ‘sun coyote’, and isil tevicnikic ‘white coyote’.” (Strong 1929, in reference to Cahuilla lore)

> “There is a place in the world above where Sun and Slo’w, Morning Star and Snilemun (The Coyote of the Sky - not the Coyote of this world) play peon.” (Blackburn 1975)

In fact, I hypothesize that Coyote was not the original mascot of the Takic Sky Moiety at all. I suspect Wildcat, Coyote, Buzzard, etc. were originally clan, not moiety, mascots, elevated in one tradition or another to the Earth Side or Sky Side moiety mascot roles. The Tongva name for Coyote is also interesting for another reason, providing additional support for an original -st- sequence in his name.

Proto-Takic *ʔísa-ta-č ‘coyote’
Tongva ʔí:-ta-r ‘coyote’
Tongva ʔí:-ta-koʔ ‘wicked’
Tongva ʔí:ša-w-t ‘wolf; greedy eater’ [coyote-big]
Cf. Tübatulabal ʔi-št ‘coyote’ (Stubbs 2011, entry 567)
Cf. Hopi ii’ist ‘coyotes’ (Stubbs 2011, entry 567)
Cf. Shoshoni isa-pai-ppï ‘coyote’ (Stubbs 2011, entry 567)

Speculating further, given the importance of Coyote as a ritual figure, the Numic and Takic branches may suggest a pattern:

Proto-Takic *ʔísa-ta-č ‘coyote-sun’
Proto-Numic *isa-pai- ‘coyote-dawn’

10 The Sky People

Just as the root *tuku may serve as an Earth moiety marker, the root *pa (a reduced form of *paha) may serve as a Sky moiety marker:

Proto-Takic *pa- ‘great; Sky moiety marker; Coyote moiety marker’
Cahuilla pa-suka-t ‘horse’ [pa ‘great’ + sukat ‘deer’]
Cahuilla pa-vu-l ‘bear shaman; bear hunter’ (Strong 1929) [pa ‘great’ + pul ‘shaman’]
Cahuilla pá-li-l ‘bat’
Cahuilla pá-yu-l ‘lizard’
Cupeño pá-xaniš ‘blackbird’
Kitanemuk pá-háŋa-č ‘a bee species, a little larger than a háŋa-č’
Kitanemuk pa-wirukuh-t ‘vulture’
Serrano pa:-wahi’ ‘wolf’
Tongva pa-í:-t ‘wood rat’
Tongva pa-í:xwoʔ ‘blackbird’
Tongva pa-kí:ša-r ‘duck hawk, prairie falcon’
Tongva pá:-ʔaša-wo-t ‘a type of bird’ [pa ‘great’ + ʔašá:wt ‘eagle’]
Tongva pa-hú:na-r ‘great bear; grizzly bear’
Tongva pá:-xava-wo-t ‘fox’
Tongva pa-šú:ka-t ‘elk’
Cf. PUA *pa- ‘big’ (entry 207)
Cf. Proto-Takic *paxáʔ ‘ritual leader; ceremonial announcer; ceremonial clown; Coyote’

11 The Dark One

Befitting his status as the ‘mascot’ of the Takic Wildcat Moiety, I would suggest that Wildcat is simply the ‘Dark One’.

Proto-Takic *tukuʔu-t ~ *tukku-t ‘wildcat; the Dark One; the Wildcat Moiety mascot’
Tongva tukú:-t ‘wildcat’
Kitanemuk tuku-t ‘wildcat’
Cf. Proto-Uto-Aztecan *tuku ‘owl’ (entry 1591)
Cf. Proto-Uto-Aztecan *tukkuC ‘wildcat’ (entry 1345)

Owl is likewise a notable member of the Earth Side, similarly named. My tentative reconstruction of a glottal stop in the Proto-Takic u-stem form of ‘wildcat’ (a type of echoic emphasis) represents my initial attempt to account for both the non-lenition of medial k and similar fortis forms in the a-stems, discussed below.

> “In connection with the wild cat as a culture hero the Southern Diegueño employ a color symbolism. They say that in the east there was a red wild cat, in the west a blue one.” (Barrett 1918)

> “[…] neophytes of the coyote moiety were painted with stripes, those of the wildcat moiety with spots […]” (Strong 1929)

On the basis of the above evidence, it is tempting to speculate that Takic culture arose from a Southern California accommodation between the local Yuman and Uto-Aztecan cultures.

12 The Conduit

In the Takic world, trees are boundary crossers, with their roots in the Earth and their crowns in the Sky. According to the Serrano, for example, the Creator died at Yahuviat, the Place of Pines (see Strong 1929), and, according to the Luiseño, Pine, Cedar, and Oak, upon their creation, were sent to the Center [of the Sacred Space] (see DuBois 1908).

> "At last a little green shoot sprouted up and grew day by day until it became the tall tree that we call now the pine, and the pine is of the same nature as the stars and holds in itself the same bright light." (Mooney 2013)

In keeping with my treatment of *tuku-ʔu-t 'Wildcat' in Section 13 and *paha-ʔa-t 'Bighorn' in Section 15, I would suggest the primary Takic representative of the Light World is *tɨva-ʔa-t 'Pinyon'.

Proto-Takic *ti-va-ʔa-t 'conifer; oak; the sacred tree; the world tree'
Cupeño té-ve-t 'conifer'
Luiseño tó:-va-ʂa-1 'valley oak'
Luiseño tó:-va-1 'pinyon tree'
Tongva to-vá:ʔa-t 'pinyon'
Kitanemuk ti-ve-t 'piñon pine'
Cf. PUA *tïvaʔ-t ~ *tïvaC 'pine' (entry 1630)

In the Takic Mourning Ceremony, the gravepole, the locus of the transition from this world to the next, is made from a carefully selected tall pine:

> "The chief feature of it was a tall painted pole called Kimul Chehenish, made very smooth from the trunk of a pine or fir, which was hung with baskets at the top to be reached by climbing as a contest of skill. This pole was as high as a house and was painted with different colors. It represented the dead man, the spirit. Different parts of the pole were painted in different colors to refer to the different parts of the body. The pole was not painted with the shape of a man, but one part of the painting meant the knee, another the arm, and so on. The top for the head was always painted white." (DuBois 1908)

> "[This kotuumot is…] painted in four colors: white, red, black, and grey." (McCawley 1998)

> "…Tomaiyowit [Earth] lay back outstretched, her feet to the north, her head to the south, her right hand to the east, her left hand to the west, and everything became quiet." (DuBois 1908)

> "Tukmishwut, the North Star, remains motionless, and all his people, the members of his 'party', move in a circle about him. This is the reason the dancing and marching are in a circle around the sacred enclosure, the fire, and so on." (DuBois 1908)

Proto-Takic ku-tumi-č 'gravepole; the North Wood'
Luiseño ku-tú:mu-t 'ceremonial painted pole'
Luiseño túm 'die (song word)'
Luiseño túm 'remember'
Luiseño tumá:mik 'to the north'
Kitanemuk kú-tomi-č ~ kú-tumɨ-č 'gravepole'
Kitanemuk tuʔmɨ-k 'be quiet'
Tongva ko- 'wood'
Tongva tomí:-n-ok 'shut up'
Tongva tomí:-na-vet-am 'feather headdress'
Tongva ko-tú:mo-t 'gravepole'
Tongva -tú:mo-n 'heel' [ceremonially pointed north]
Tongva tú:me 'north'
Cf. PNUA tïmïnïmïn 'north, west' (entry 1544)

Cf. PNUA *kuttumu ‘pole of solemn purpose’ (entry 2411)
Cf. Chumash sq’oq’om ‘gravepole, ideally made of red pine; the road to heaven’

13 The Bright One

I have suggested Coyote was not the original mascot of the Takic Sky Moiety, but an interloper. He may have replaced Bighorn in this regard.

Proto-Takic *pa(h)a-ʔa-t ‘bighorn’ (as suggested by a Tübatulabal cognate) may be a fortis version of *paha ‘bright’ (emphasizing his ritual role) just as Proto-Takic *tu(k)kuC-t ‘wildcat’ may be a fortis version of *tuku ‘dark’ (emphasizing hers). Put more simply, Bighorn may be ‘The Bright One’ just as Wildcat is ‘The Dark One’. Befitting a member of the Sky Side, Bighorn lives inland (in the desert mountains), is diurnal (becoming active at dawn), and is among the most common subjects of Southwestern rock art (see Valdez & Krausman 1999).

> “We believe that Newberry Cave was a site for an exclusively male, bighorn sheep, totemic, hunting society (moiety or clan) where rituals were carried out to promote the life and health of a key supernatural, ancestral, totemic animal - the desert bighorn sheep.” (Garfinkel 2015)

> “In Numic ideology, the bighorn appears to have represented the entire category of large game animals and was closely related to the hunt, the killing of big game, and the provisioning of meat that evinced heightened status and male hunting prowess. These animals also exhibited close associations with the concepts of prestige, power, and strength, and acted as referents to the men’s coming-of-age ceremony and sexual maturity.” (Garfinkel 2015)

Beyond Takic and Numic, consider Hopi paalata ‘to shine a light’, paaho ‘prayer stick’, Panwu Kachina ‘The Bighorn Kachina’, and pahaana (Kroeber baxana) ‘the lost brother, an Anglo’. The latter, evidently an analogue of Queztalcoatl, is a god of the East (the Sky Side). Legend has it that, when he returns, he will bring back the Fire Clan’s portion of a sacred stone, and that he will come wearing red (see Courlander 1971).

14 The First Woman

Of course, in the Takic belief system, Animals are not the only People. Natural Phenomena, Plants, and, of course, Human Beings are People too. There is some evidence for the concept of The First Woman, parallel to the better documented concept of The First Man.

Tongva Tobohar ‘The First Woman’ (Kroeber 1925)
Tongva tokó:-r ‘woman’
Tongva tokú:-r-aw-t ‘pretty’
Kitanemuk tuhu ~ tuqu ‘a man who lives like a woman’

Perhaps the Tongva form tokó:-r ‘woman’ was specifically associated with a Wildcat Women (i.e. a woman born into the Wildcat Side), but in any event another set appears to reinforce the connection between women and the Earth Moiety:

Proto-Takic ʔoka-č ‘earth; sand; Earth Woman’
Kitanemuk ʔoka-č ‘sand; sandy area’
Tongva Auzar ‘The First Woman; the Mother of Chinigchinich’ (Boscana 1846)
Tongva ʔó:xo-r ‘earth; land; world’
Tongva -ó:xo ‘elder sister’
Fernandeño ö:xa-r ‘land’ (Stubbs 2011)
Cf. PUA *oka ‘sand; earth; rock’ (entry 1830)
“During this year the girl was prohibited from eating either meat or salt. At its close other rites were performed, and a lecture or counsel was given to the girl on much the same lines as that given to the boys. She was cautioned against being stingy, against dissembling, and against

looking sidewise. She was also told not to eat jackrabbit or venison. After this lecture the girl was freed from all restrictions." (Sparkman 1908)

15 The Shaman

In this analysis, shamans serve as mediators between worlds, performing the rituals that make such dangerous transitions possible:

"When you die your spirit will rise to the sky, and people will blow (three times) and will make rise your spirit." (Sparkman 1908)

Proto-Takic *púha-ča 'shaman; a member of the elite; the blowing one'
Cahuilla pú:-l 'witch doctor, medicine man'
Cahuilla pa-vu-l 'bear shaman; bear hunter' (Strong p. 115)
Cahuilla pú:-wax 'to perform a witch dance'
Luiseño pú:-la 'shaman'
Luiseño pú-mmal-um 'male initiates'
Luiseño pú:-lu 'to do the púulush dance'
Luiseño pú:-lu-š 'a shaman's magic power'
Luiseño pú:-mu-š 'the ceremonial eagle feather worn on the head'
Luiseño pu-ki' 'dandelion'
Luiseño pú:-ma-l 'an initiated boy, after drinking the jimsonweed drink'
Tongva pú:-ʔe 'to blow'
Tongva pú:-ra 'Chinigchinich' (DuBois 1908, "in the old language of the coast")
Kitanemuk puh-ea? 'to blow on'
Kitanemuk puh-k 'to spray water'
Cf. Proto-Takic *pú:hi-wC-t 'roadrunner; the First Man's doctor'
Cf. PUA *puca 'to blow' (entry 261)

16 The First Man

The pivotal figure of the Gabrielino/Luiseño religious complex is Woyoot/Wiyot, both the wisest of men and the first to leave this world.

Proto-Takic *wi(ʔ)-wi(ʔ)-wïC-t 'the wise one; the first man'
Luiseño wí-s-ku-n 'chipmunk'
Luiseño wí-š-ku-š 'brain'
Luiseño wí-t 'an interjection used to announce the presence of a powerful animal spirit'
Luiseño wuyó:-t 'religious character; Father Moon'
Tongva we-wy-ó:-t ~ wo-y-ó:-t 'a mythic figure'
Tongva wí:-y 'acorn mush'
Tongva -wí:-we-n 'root'
Fernandeño Veat 'Woyoot; the first man to die' (Boscana 1846)
Fernandeño Ouiamot 'Woyoot's successor, Chinigchinich' (Boscana 1846)
Kitanemuk wiʔ-wi(ʔ) 'brains'
Chumash wi-wi-y-i-t (Hudson & Blackburn 1978) [borrowed from a Takic language]

The semantics of this set are difficult, but sortable, I think. Brains may be actually or metaphorically consumed as a means to acquire wisdom. It was Wuyoot/Woyoot who gave mankind the gift of acorns. An acorn is like a brain, in that all the good stuff is contained within a hard shell. Chipmunks store and eat acorns. Chipmunk was the one who crafted Wiyot's coffin from a tree. Acorns produce roots. Wiyot is at the root of humankind's family tree.

"There was a man among them named Wískun (now a tiny squirrel), and when Ouiot was burned, he stood up and addressed the people; and he called the clouds from the mountains to come, and the clouds and fog from the sea to gather and fall in showers upon the earth to blot out all the

> tracks that Ouiot had made when he moved about upon the earth, so that nothing could be seen." (DuBois 1908)

The Kitanemuk and Chumash evidence may be decisive here. The Chumash form was presumably borrowed from an earlier Takic (Kitanemuk, Tataviam, or Tongva) form and supports the reconstruction of a reduplicated, augmented proto-form with the meaning *wi(ʔ)-wi(ʔ)-wiC-t literally meaning 'big brains'. Note the similarity of Cahuilla sísi-l^{y} 'chipmunk' to Tongva -sesoʔ 'brains', the latter of which is presumably borrowed from Spanish sesos 'brains'. Woyoot was succeeded by Chinigchinich, a mysterious culture hero who has been analyzed every which way. In my own view, the Gabrielino/Luiseño pair of *Wiwiyɨt and Chinigchinich (etymology unknown) correspond to the creator brothers of the Cahuilla, Cupeno, and Serrano traditions.

References

Agua Caliente Cultural Museum. 2017. In The Beginning. Manuscript.

Anderton, Alice Jeanne. 1988. The Language of the Kitanemuks of California. UCLA Dissertations.

Applegate, Richard. 1979. The Black, The Red, and the White: Duality and Unity in the Luiseño Cosmos. The Journal of California and Great Basin Anthropology.

Bean, Lowell John. 1974. Mukat's People: The Cahuilla Indians of Southern California. University of California Press.

Blackburn, Thomas C. 1975. December's Child: A Book of Chumash Oral Narratives. University of California Press.

Blust, Robert and Trussell, Stephen. 2017. The Austronesian Comparative Dictionary. Web Edition.

Boscana, Geronimo, translated by Alfred Robinson. 1933. Chinigchinich.

Courlander, Harold. 1971. The Fourth World of the Hopis. University of Mexico Press.

da Matta, Roberto. 1982. A Divided World: Apinayé Social Structure. Harvard University Press.

Du Bois, Constance Goddard. 1908. The Religion of the Luiseño Indians of Southern California. Dodo Press.

Elliott, Eric Bryant. 1999. Dictionary of Rincon Luisefio. UCSD Dissertations.

Garfinkel, Alan P. et al. 2015. Sacred Stories and Symbolic Songs of the Immortal Bighorn: Newberry Cave. Manuscript.

Garfinkel, Alan P. et al. 2016. Ritual, Ceremony and Symbolism of Archaic Bighorn Hunters of the Eastern Mojave Desert: Newberry Cave, California. Rock Art Research, Vol. 33, No. 2, pp. 193-208.

Gifford, Edward Winslow. 1918. Clans and Moieties in Southern California. Univeristy of California Publications in American Archaeology and Ethnology, Vol. 14, No. 2.

Hill, Jane and Velasquez, Rosinda. 1973. Mulu'wetam the First People: Cupeño Oral History and Language. Malki Museum Press-Ballena Press.

Hudson, Travis and Blackburn, Thomas. 1978. The Integration of Myth and Ritual in South-Central California: The Northern Complex. The Journal of California Anthropology.

Hunter, John M. 1973. Geophagy in Africa and in the United States: A Culture-Nutrition Hypothesis. Geographical Review, Vol. 63, No. 2, pp. 170-195.

Hyde, Villiana Calac and Elliot, Eric. 1994. Yumáyk Yumáyk. University of California Press.

Kehoe, Alice Beck. 2002. America Before the European Invasions. Routledge.

Kroeber, Alfred Lewis. 1925. Bureau of American Ethnology of the Smithsonian Institution, Bulletin 78.

Lévi-Strauss, Claude, translated by John Weightman and Doreen Weightman. 1983. The Raw and the Cooked. University of Chicago Press.

McCawley, William. 1996. The First Angelinos: The Gabrielino Indians of Los Angeles. Malki Museum Press-Ballena Press.

Mallory, J.P. and Adams, D.Q. 2006. The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World. Oxford University Press.

Malotki, Ekkehart. 1983. Hopi Time: A Linguistic Analysis of the Temporal Concepts in the Hopi Language. Mouton Publishers.

Mamet, Ingo. 2010. The Luiseño Absolutive Suffix: Diachronic Perspectives. Anthropological Linguistics, Vol. 52, No. 3-4, Fall and Winter, pp. 239-273.

Mooney, James. 2013. History, Myths, and Sacred Formulas of the Cherokees. US Bureau of American Ethnology.

Munro, Pamela, with input from the Gabrielino-Tongva Language Committee. 2017. Gabrielino/Tongva/Fernandeño Vocabulary with an English Index. Manuscript.

Munro, Pamela. 2016. Evidence about Tongva Religious Figures. Manuscript.

Puhvel, Jaan. 1989. Comparative Mythology. John Hopkins University Press.

Scarcliff, Brent. 2017. A Proto-Takic Etymological Dictionary. Manuscript.

Seiler, Hansjakob Seiler and Hioki, Kojiro. 1979. Cahuilla Dictionary. Malki Museum Press.

Serrano Language Revitalization Project. 2012. Serrano Words. San Miguel Education Department.

Sparkman, Philip Stedman. 1908. The Culture of the Luiseño Indians. Univeristy of California Publications in American Archaeology and Ethnology, Vol. 8, No. 4, pp 187-234.

Strong, William Duncan. 1929. Aboriginal Society in Southern California. Univeristy of California Publications in American Archaeology and Ethnology, Vol. 26.

Stubbs, Brian D. 2011. Uto-Aztecan: A Comparative Vocabulary. Rocky Mountain Books and Productions.

Wang, Robin R. 2012. Yinyang: The Way of Heaven and Earth in Chinese Thought and Culture. Cambridge University Press.

Constraints on Verbs in Series

Ronald P. Schaefer and Francis Egbokhare
Southern Illinois University Edwardsville and University of Ibadan

1 Introduction

Verbs in series continue to attract linguistic attention as a syntactic phenomenon (Foley and Olson 1985, Sebba 1987, Baker 1989, Watters 2000, Crowley 2002, Aikhenvald and Dixon 2006) and as one form of complex predicate (Amberber, Baker and Harvey 2014). Generally ignored in much of this discussion is the nature and configuration of meaning components for verbs in series. As a consequence, semantic restrictions on verb co-occurrence within serial verb constructions remain largely under described.

For this paper, we consider both syntactic and semantic properties of verbs in series in order to highlight construction constraints. Our results are illustrated using southern Nigeria's Emai (Edoid/West Benue Congo, Williamson and Blench 2000). Relatively strict SVO, Emai exhibits lexical and grammatical tone, little segmental inflection and few prepositions. Our analysis incorporates data gathered as a result of text collection (Schaefer and Egbokhare 1999), dictionary construction (Schaefer and Egbokhare 2007) and reference grammar description (Schaefer and Egbokhare 2017).

Our illustrations are consistent with classic serial verb properties (Aikhenvald and Dixon 2006): a verb sequence sharing tense, aspect and polarity under a single intonation contour and a single predicate with no overt marking of syntactic dependency between verbs.

Emai manifests a robust system of serial verb constructions whose internal relations reveal restrictions on order and co-occurrence. To characterize these relations among intransitive verbs, we utilize semantic constructs articulated in association with Sorace's (2000, 2004) aspectual/thematic (AT) hierarchy. For transitive verbs, we rely on Levin and Rappaport Hovav's (1995, 2005, 2010) notion of manner/result complementarity. Both call attention to a basic distinction between two verb types: manner/process vs result/transition.

2 Constraints on Intransitive verbs

Sorace (2000, 2004) developed an aspectual thematic (AT) hierarchy to account for auxiliary variation (BE ~ HAVE) among perfective constructions in Standard Average European. She unified two hierarchic schemas. An aspectual or transition hierarchy reflecting degrees of telicity is illustrated by the English verbs *be>remain>rot>arrive*. A thematic or process hierarchy displaying degrees of agentivity is shown by the verbs *work>run>shiver*. Linking these two are verbs like *melt* that share properties of each hierarchy.

Sorace's unified AT hierarchy incorporates the transition and process sub-hierarchies into one:

CNM	COM	UCA	ATS	EXS	COS	CHS	CLO
work	run	shiver	melt	be	remain	rot	arrive

CNM=controlled non-motion,
COM=controlled motion,
UCA=uncontrolled activity,
ATS=antitransitive state
EXS=existence state,
COS=continuative state,
CHS=change of state
CLO=change of location

* Data for this paper derive from research sponsored by the National Science Foundation, (BNS #9011338 and SBR #9409552), the Department of State (College and University Affiliations Program grant ASJY 1333) and the National Endowment for the Humanities (PD-50004-06). While we thank these institutions for their support, only we are responsible for data interpretation.

Proceedings of *WECOL 2017*

We now consider Emai intransitive verbs in serial verb constructions. Our aim is to assess how the AT hierarchy may enhance our understanding of constraints governing intransitive verbs occupying positions in series.

Relative to the AT hierarchy, Emai verbs in series exhibit several constraints. The principal class constraint is process verbs precede transition verbs. Controlled non-motion *gua* 'heap' precedes change of location *vade* 'come' (1), and controlled motion *la* 'run' precedes change of location *vare* 'come' (2).[1]

(1) *òjè ò ó gùà vádé.*
Oje SC C heap come
'Oje is heaping (yams) and coming.'

(2) *òjè lá váré vbì ìwè.*
Oje PRP.run come LOC house
'Oje ran to the house. / Oje came running to the house.'

Verb transposition would be ungrammatical.

(3) **òjè vádé gùà.*
Oje coming heap
'Oje is coming and heaping (yams).'

Second, process verbs regardless of order never co-occur in series. Controlled non-motion *ta* 'talk' and controlled motion *gbe* 'dance' are unacceptable in series (4). Similarly, controlled motion *la* 'run' and controlled non-motion *je* 'laugh,' as verbs in series (5), are ungrammatical.

(4) **élí ívbèkhàn ò ó tà gbé vbì òsíé'.*
the youths SC C talk dance LOC entertainment
'The youths are talking and dancing at the event.'

(5) **òjè ò ó là jé.*
Oje SC C run laugh
'Oje is running and laughing.'

The third constraint bears on the order of intransitive transition verbs. Some transition types co-occur. For instance, change of state *daa* 'raise' precedes existence state *muzan* 'stand.'

(6) *òjè dáá múzán.*
Oje PRP.raise stand
'Oje stood up.'

As well, transition verb co-occurrence is permissible for change of location. The verb *o* 'enter' can precede verb *raale* 'leave.'

(7) *òjè ó vbí ékéín ímè ráálè.*
Oje PRP.enter LOC inside farm leave
'Oje entered the farm and left.'

However, verbs of change of location never precede any other transition verb type. For example, change of location *re* 'arrive' fails to precede existence state *dia* 'sit.'

[1] Orthographic conventions for Emai are consistent with those in Schaefer and Egbokhare (1999, 2007, 2017), where **ọ** represents a lax mid back vowel, **ẹ** a lax mid front vowel, and **vb** a voiced bilabial approximant. With respect to tone, acute accent marks high, grave accent signals low, and acute accent followed by an apostrophe designates high downstep. Abbreviations for grammatical morphemes used throughout this paper include: C=continuous, LOC=locative, PRP=present perfect, SC=subject concord.

(8) **òjè ré díá vbì àgá.*
Oje PRP.arrive sit LOC chair
'Oje arrived and sat in the chair.'

3 Constraints on transitive verbs

Co-occurrence constraints of a like nature affect Emai transitive verbs in series. To characterize these, we employ the Levin and Rappaport Hovav (2010) constructs Manner and Result.

Result verbs characterize event change quite differently from Manner verbs (Beavers 2013). Result forms express scalar change. A scale is a set of degrees indicating values of measurement on a particular dimension along with an associated ordering relation for those values. Result verbs convey scalar change along a property or path, or of volume or existence. Examples include English *break*, *cool*, *enter*, *build*, *eat* and *drink.*

Manner verbs express non-scalar change. They convey complex changes with no privileged scale of change. They represent a complex sequence of separate changes that collectively define an action. They do not add up to a single cumulative change along any single dimension. Examples include English *run*, *scrub*, and *sweep.*

These two verb types, Manner and Result, are frequent partners in a single serial verb construction. They are also found in other complex predicate types, such as resultative constructions that exist in some languages. English, for instance, shows the resultative complex predicates *sweep clean* and *smash flat*, where *sweep* and *smash* represent Manner, while *clean* and *flat* characterize Result.

In Emai, verb series constructions show sequences of transitive Manner and Result forms. These sequences, however, are highly constrained. Transitive Manner *sua* 'push' can precede transitive Result *ye* 'move toward,' as in (9). But Result *ye* can never precede Manner *sua* (10).

(9) *òjè súá ìmátò yé èkó.*
Oje PRP.push car move.toward Lagos
'Oje has pushed the car toward Lagos.'

(10) **òjè yé èkó súá ìmátò.*
Oje PRP.move.toward Lagos push car
'Oje moved toward Lagos and pushed the car.'

Transitive Manner verbs in series in Emai are highly constrained. A Manner verb like *laa* 'carry in turns, take turns carrying' followed by another Manner verb, e.g. *la* 'chew,' is unacceptable in a serial verb construction.

(11) **yàn á làà éánmí lù.*
they C carry.in.turn meat chew
'They took turns carrying meat and chewing it.'

Result verbs, on the other hand, do co-occur in series. Such is the case with transitive Result verbs of location change. In (12), *shan* 'move through' and its direct object co-occur with *ye* 'move toward' and its object.

(12) *òjè shán égbóà yé ìwè.*
Oje PRP.move.through backyard move.toward house
'Oje moved through the backyard toward the house.'

Summing up for Emai, we have seen in the preceding examples three basic properties of Manner and Result transitive verbs. First, Manner and Result exemplars co-occur in verb series (9); second, Manner and Manner verbs fail to combine (11); and third, Result and Result associate in a serial sequence (12).

We direct attention now to some problematic facets of other serial verb constructions vis-à-vis the constructs Manner and Result. In (13), Manner *de* 'buy' combines with Result *e* 'eat,' the latter a verb characterizing incremental change along a dimension of consumption. However, in 14 and 15, respectively, we see that each of the Manner verbs *nwu* 'carry' and *lie* 'collect' fail to co-occur with *e* 'eat.'

(13) *òjè dé émà é.*
Oje PRP.buy yam eat
'Oje bought yam and ate it.'

(14) **òjè nwú émà é.*
Oje PRP.carry yam eat
'Oje carried yam and ate it.'

(15) **òjè lie ituu é.*
Oje PRP.collect mushroom eat
'Oje collected mushrooms and ate them.'

In response to proposed serial verb examples like (14) and (15), we consistently receive feedback that the subevents represented by the separate verbs are difficult to relate contextually. They lack a sufficiently close association, apparently one of purpose or intention. That is, buying something can be viewed as intentionally or purposively related to eating that something. But neither carrying nor collecting something can be so viewed relative to eating of the same. This lack of association would suggest that intentionality must play some role in selecting verbs for at least the type of verb series construction shown in (13).

Co-participant verbs in series present a different kind of problem. They are found in a serial verb construction where one verb phrase aligned with an event co-participant precedes another verb phrase that conveys the event core, at least for purposes of event translation. Verb series constructions of this type display a strict precedence relation.

There are in Emai three co-participant constructions. In (16), the verb complex *de baa* 'join' and its subject co-participant precede the verb *sua* 'push.' In (17), the verb *kpaye* 'accompany' and its subject co-participant precede *ta* 'speak.' And in (18), the verb *kpaye* 'replace, take up a place on behalf of' and its subject co-participant precede verb *e* 'eat.'

(16) *ólí ómòhè dé báá élí ívbèkhàn súá ìmátò.*
the man PRP.reach join the youths push car
'The man joined the youths to push the car / pushed the car with the youths.'

(17) *ólí ómòhè ò ó kpàyè ójé tà étà.*
the man SC C accompany Oje speak word
'The man is speaking with Oje.'

(18) *ólí ómòhè kpáyé òlólò é ólí émàè.*
the man PRP.replace Ololo eat the food
'The man took Ololo's place and ate the food / ate the food on behalf of Ololo.'

How are the verbs in these constructions assigned to the categories Manner and Result? We take the first example with *de baa* 'join' and *sua* 'push' to develop an initial impression of the dilemma that arises. We could view *de baa* as a realization of Manner in construction with *sua*, which we previously saw in (9) behaved like a Manner verb. But this would mean that a verb series could encode a Manner + Manner sequence, in violation of the constraint that was evident previously with serial constructions comprised of verbs that were intransitive (5) or transitive (11).

Alternatively, we could view *de baa* as a realization of Result in series with *sua*. Recall, however, that previously *sua* has behaved as a Manner verb. Such a Result + Manner pairing of verbs would contradict the precedence relation in Emai between Manner and Result, which characterized verb series combinations that were transitive (10) or intransitive (3). Assuming Result+Result or Manner+Result for this verb series is not feasible either, since both would require that *sua* exemplify Result. Examples like (9), which we reviewed prior, indicate that *sua* cannot be so identified.

None of the alternatives open to us seems to characterize consistently co-participant verbs as an exponent of Manner or Result. Perhaps there is something about the function of co-participant constructions that is being ignored. All identify a situation where the co-participant assumes a location associated with an argument of the core verb. Co-participant joiner co-locates with the joinee.

This function suggests that there may be a connection between the co-participant relation and locatives more generally. Emai, like other languages, distinguishes between locative arguments and locative adjuncts, or between inner locatives and outer locatives. Syntactically coding this distinction in Emai is the verb *za* and its precedence relation relative to another verb in series. This coding difference becomes evident when we compare the shape of canonical constructions to their non-canonical counterparts such as contrastive focus and content interrogatives (Schaefer and Egbokhare 2014).

In canonical constructions, both locative types, adjunct in (19a) vs argument in (19b), are syntactically marked by the postverbal preposition *vbi.*

(19) a. *ólí ómòhè gbé ólí ófé vbí ímè.*
the man PRP.kill the rat LOC farm
'The man has killed the rat on the farm.'
b. *ólí ómòhè ó vbì ìwè.*
the man PRP.enter LOC house
'The man has entered the house.'

In non-canonical constructions, whether contrastive focus or interrogative, a locative adjunct requires that its predication include the verb *za* in series. In addition, *za* must precede the core verb found in the canonical predicate. Without *za*, a locative adjunct in either focus or interrogative position would be ungrammatical.

(20) a. *ímè lí ólí ómóhé zá' gbé ófè.*
farm PF the man PAP.move.loc kill rat
'It was on the farm that he killed a rat.'
b. *ébé' ólí ómóhé zá' gbé ófè?*
where the man PAP.be.loc kill rat
'Where did the man kill a rat?'

In contrast, a locative argument in a non-canonical construction disallows the verb *za* in series.

(21) a. *ólí íwé nà lí ólí ómóhé ó'-ì / *zá ó'-ì.*
the house this PF the man PAP.enter-F move.loc PAP.enter-F
'It was this house that the man entered.'
b. *ébé' ólí ómóhé ó'-ì? / *zá) ó'-ì?*
where the man PAP.enter-F move.loc PAP.enter-F
'Where did the man enter?'

In other constructions, *za* designates a source argument relative to a goal. Nonetheless, *za* never occurs as sole verb in a simple predication.

(22) a. *òjè zá vbì áfúzé' shán sé vbí òkè.*
Oje PRP.move.loc LOC Afuze walk move.as.far.as LOC Oke
'Oje walked from Afuze to Oke.'
b. *òjè zá vbí ímè ráálè / *zá vbí ímè.*
Oje PRP.move.loc LOC farm move.away PRP.move.loc LOC farm
'Oje moved away from the farm.'

It is the verb *za* in series and its precedence relation that appears relevant to co-participant constructions. In both, we have a verb in a precedence relation relative to another verb in series. In neither instance, does the argument associated with the verb observing the precedence relation serve as an argument of the core verb. Noun phrases associated with co-participant verbs and the verb *za* are outliers; they are external to the core verb. In short, both are adjuncts. As such, co-participant verbs and *za* are not governed by the same constraints as those governing core verbs. We would like to propose, therefore, that verbs coding co-participants are neither Manner nor Result exponents. Instead, they appear to realize a third verb type, grammatical adjunct for the moment, that has a syntactic function not unlike other adjunct marking verbs such as *za*. Neither introduces an event participant that serves as argument of a core verb.

4 Conclusion

In the preceding, we examined constraints affecting transitive and intransitive verbs in series in the Edoid language Emai. We identified verb combinations that were admissible as well as those that were not. The resulting constraints were framed in terms of the verb types Manner (process) and Result (transition). Verb series predications coding an event co-participant, however, challenged these constraints. Common to their predications was a co-locate function for the co-participant relative to the core verb participants. Based on this function, we examined Emai locatives in canonical and non-canonical structures. In non-canonical focus and interrogative constructions, Emai syntactically distinguished between locative adjuncts and locative arguments. Locative adjuncts were coded by verb *za* and its precedence relation relative to another verb in series. In several respects, these syntactic conditions mirrored those of co-participant predications. As a consequence, we proposed that noun phrases associated with co-participant verbs were adjuncts and that the verbs themselves served a function related to the grammatical marking of adjuncts. We also noted that verbs marking co-participants were exponents of neither Manner nor Result. Instead, we suggested that they reflected a third verb type that we, for the moment, identified as grammatical adjunct. No verbs of this third type introduce an event participant that serves as a predication argument. We are hopeful that future attention to other serial verb predications may clarify this initial hypothesis and provide needed insight into the internal structure of serial verb constructions.

References

Aikhenvald, Alexandra Y. 2006. Serial verb constructions in typological perspective. In A.Y. Aikhenvald & R.M.W. Dixon (eds.), *Serial verb constructions: A cross-linguistic typology*. 1-68. New York: Oxford University Press.

Amberber, Mangistu, Brett Baker & Mark Harvey. 2014. *Complex predicates: Cross-linguistic perspectives on event structure*. New York: Cambridge University Press.

Baker, Mark. 1989. Object sharing and projection in serial verb constructions. *Linguistic Inquiry* 20.4.513-553.

Beavers, John. 2013. Aspectual classes and scales of change. *Linguistics* 51.4. 681-706.

Crowley, Terry. 2002. *Serial verbs in Oceanic: A descriptive typology*. New York: Oxford University Press.

Elugbe, Ben. 1989. *Comparative Edoid: Phonology and lexicon*. Port Harcourt: University of Port Harcourt Press.

Foley, William A. & Mike Olson. 1985. Clausehood and verb serialization. In Johanna Nichols and Anthony Woodbury (eds.), *Grammar inside and outside the clause*. 17-60. New York: Cambridge University Press.

Levin, Beth & Malka Rappaport Hovav. 1995. *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge, MA: The MIT Press.

Rappaport Hovav, Malka & Beth Levin. 2010. Reflections on manner/result complementarity. In Malka Rappaport Hovav, Edit Doron & Ivy Sichel (eds.), *Lexical semantics, syntax and event structure*. 21-38. New York: Oxford University Press.

Schaefer, Ronald P. and Francis O. Egbokhare. 1999. *Oral tradition narratives of the Emai people, parts I and II*. Hamburg: LIT Verlag.

Schaefer, Ronald P. & Francis O. Egbokhare. 2007. *A dictionary of Emai: An Edoid language of Nigeria*. Köln: Rüdiger Köppe Verlag.

Schaefer, Ronald P. & Francis O. Egbokhare. 2014. Emai's variable coding of adjuncts. *Linguistic Discovery* [special issue on arguments and adjuncts cross-linguistically, ed. by Soren Weichmann] 12.2: 12-26.

Schaefer, Ronald P. & Francis O. Egbokhare. 2017. *An Emai grammar*. [Mouton Grammar Library Series] Berlin: Mouton de Gruyter.

Sebba, Mark. 1987. *The syntax of serial verbs*. Amsterdam: John Benjamins.

Sorace, Antonella. 2000. Gradients in auxiliary selection with intransitive verbs. *Language* 7 (4): 859-890.

Sorace, Antonella. 2004. Gradience at the lexicon-syntax interface: Evidence from auxiliary selection and implications for unaccusativity. In Artemis Alexiandou, Elena Anagnostopoulou & Martin Everaert (eds.), *The unaccusativity puzzl*e. 243-268. New York: Oxford University Press.

Watters, John R. 2000. Syntax. In Bernd Heine & Derek Nurse (eds.), *African languages: An introduction*. 194-230. New York: Cambridge University Press.

Williamson, Kay & Roger Blench. 2000. Niger Congo. In Bernd Heine & Derek Nurse (eds.), *African languages: An introduction*. 11-42. New York: Cambridge University Press.

The Changing of Arabic Terminology in Times of War and Displacement*

Jamillah Sleiman and Mythili Menon

Wichita State University

1 Introduction

Refugee crisis or massive forced population movements are conflict driven displacements often from rural to urban areas. In 2017, around 65.6 million people were displaced due to conflicts, according to a report by the United Nations Development Program. The history of the Middle East North Africa region (MENA), in particular, has been shaped by episodes of forced migration in the post-World War II era. The partition of Palestine in 1947 and the subsequent war in 1948 created one of the longest-lasting refugee waves. More than 700,000 Palestinians were displaced to refugee camps in Jordan, Lebanon, and Syria and were then forced to displace again due to the 1967 June war. Many families who migrated to Lebanon and Kuwait had to evacuate their homes again, due to the Lebanese Civil War from 1975-1991, and later the 1990 Iraqi invasion of Kuwait which led to the Kuwait War from 1990-1992. Thus, Palestinian refugees have undergone several waves of forced displacement due to war, political and religious conflicts.

The official language of the state of Palestine is Arabic. This is also the case in the countries where the first refugee camps were set up, i.e. Jordan, Lebanon, and Syria. However, Arabic exhibits diglossia, a linguistic situation where two varieties of the same language exist to fulfil different social functions. Very often in Arabic-speaking communities, there are two varieties of Arabic used: Modern Standard Arabic (High variety) and the various, regional colloquial Arabic (Low variety), such as Egyptian Arabic, Moroccan Arabic. Some of these regional varieties are mutually unintelligible with each other, for example, the variety spoken in Algeria is linguistically different from the variety spoken in Egypt. The Eastern dialects exist on a continuum, called the Levantine continuum and cross-dialect communication is easier. This continuum cuts across Syria, Jordan, Lebanon, and Palestine. At the same time, the regional varieties, such as Lebanese Arabic and Syrian Arabic remain indigenous and spoken primarily in that particular country and region. Previous studies have looked at the phonetic and phonological changes existing in a particular variety when compared to another variety, but no study has looked at the change of Arabic across the Levantine continuum through the eyes of war and displacement.

This paper traces the development and changes of the Arabic language through the journey of a Palestinian family as they find refuge in a variety of countries. Through a variety of online resources and first-hand accounts from family members, this paper sheds light on the tribulations that this family has faced, and how these conflicts have influenced the way they speak Arabic to this day. The family has resided and had children in a variety of countries including Palestine, Lebanon, Kuwait, and now America. Although the general Arabic spoken within the family is the same, there are underlying differences in the pronunciation of words. The specific goal of this paper is to show the impact of the Palestinian-Israeli conflict on the way this family speaks Arabic today in America. The Palestinian-Israeli Conflict is fueled by social, political, historical, ethnic, and religious elements that has created

* We thank Khadijah Salah and Riad Sleiman for the data presented in this study. We also thank Francis X Connor, Aaron Rodriguez and other members of the Linguistics Club at Wichita State, and the audience at WECOL for helpful comments and suggestions.

Proceedings of *WECOL 2017*

tensions among the people in these lands for many years, but specifically since 1948 when Israel was officially recognized as a nation (García, 2014). This conflict is important to this paper because it highlights the internal struggle an elder in the family, namely the grandmother Khadijah, experienced during the time of being displaced from her home, and causing her displacement to other lands.

The rest of the paper is structured as follows. We begin with detailing the Israel Palestine conflict in Section 2, followed by Section 3 on the different varieties of Arabic. We then introduce the participants of the study in Section 4, and then move onto the data collection in Section 5 and the results of the study in Section 6. We conclude with a discussion and some plans for further studies.

2 The Israel Palestine Conflict

The Israel Palestine conflict, occurred over the course of many years, can be said to have begun early in the 20th century and it is still ongoing today. It started when the Jewish settlers began settling in Palestinian communities and resistance began: "By the beginning of the 20th century, with the spread of the spirit of nationalism to the area, they began to think and talk about 'decentralizing' Ottoman rule....The spread of Jewish settlement in Palestine resulted in friction between neighboring Arab and Jewish communities" (Morris 2004). This friction created a lot of separation within these communities. Britain took control over Palestine, but their control meant a split commitment between the Jewish settlers and the Arab inhabitants this "became apparent and the inevitable clash...became manifest" (Morris, 2004). This clash between the two resulted in many people losing their homes or their lives. The mother Khadijah, who was interviewed throughout this research, recalls these times with much sadness because she left her home in Tiberias, Palestine and all her belongings behind. "For the Arabs— and in particular the Palestinians— the events of 1948 are... catastrophe, trauma, and disaster. For the Jews—and in particular the Israelis—the war was a war of independence and 1948 for them was a year of miraculous and glorious events" (Pappe, 2006). These feelings of "trauma and disaster" are similar feelings many refugees have when they are forced from their countries in pursuit of a safer life.

"It is clear that the first 75,000 Arabs who left their homes in Palestine by February March 1948—- before Jewish victories of April... Had these 75,000 not taken flight, Palestinian and Middle Eastern history would have taken an entirely different course" (Teveth, 2006). This resulted in many Palestinians traveling to neighboring countries to seek refuge and safety. The family in this investigation left Palestine in 1948 for a better life away from the politics and conflict. "Following World War II, amid outrage at the atrocities of Nazi Germany's Final Solution and the Holocaust, Zionism gained broader international appeal across Europe and North America as sympathy for the Jewish plight increased. In a resolution approved in November 1947, the United Nations partitioned the territory of Palestine between Arabs and Jews, allowing for the modern nation of Israel to be formed as an independent Jewish state; Israel was officially born on May 14, 1948. This move infuriated much of the Arab and Islamic worlds." (García, 2014). With this one political move, this land was no longer Palestinian land but now Israeli land. To this day, this land is recognized as the nation of Israel. Currently, there are millions of Palestinians who now live in other countries across the world. Palestinian refugees and descendants can be found across the middle east and even in many parts of the U.S.

Figure 1: Visual Representation of the Loss of Palestinian Land 1947 to Present.

As the figure shows, there is a drastic difference in what area forms Palestine in the 1947 before the partition and after the partition. Currently 2.9 billion Palestinians live in the West Bank (the green belt near the coast, in Figure 1 (present)), and 1.8 million in the Gaza strip.

2.1 *The Setting of the Study* All of this conflict led the focus family of this research to move from Tiberias, Palestine to Baalbeck, Lebanon. Khadijah Salah, the grandmother mentioned throughout the study, left Tiberias, Palestine with her husband, Najib (Al Keylani) Sleiman, in 1948. They relocated to Baalbeck, Lebanon. They resided there with their 5 kids. It was in Baalbeck, Lebanon that Riad Sleiman was born in 1964. After 20 years of residing in Lebanon, Najib Sleiman decided to move the family to Khaitan, Kuwait for further financial opportunities. Riad decided to move to Wichita, Kansas to pursue an engineering degree in Aerospace and eventually brought his parents to America to live with him. Najib Sleiman, the grandfather, passed away on June 26, 2002 in America.

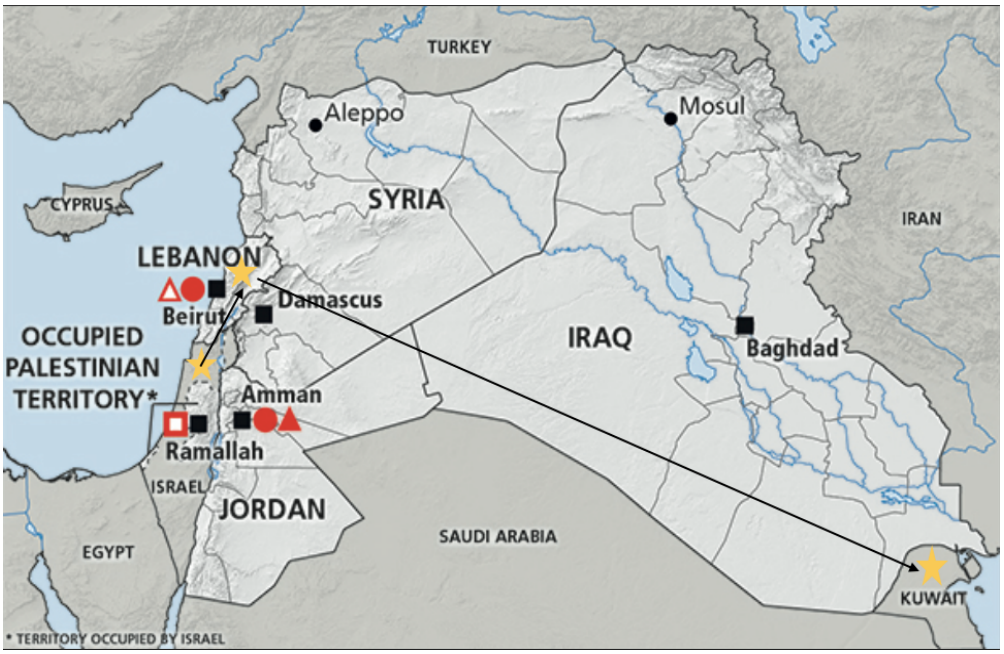


Figure 2: Map of the movement of the Sleiman family between 1948-1964.

3 Varieties of Arabic

The Arabic language encompasses a variety of dialects from multiple countries: "Those varieties include one "written" form, *Modern Standard Arabic* (MSA), and many "spoken" forms, each which is a regional dialect" (Zaidan and Callison-Burch, 2014). This written Arabic known as MSA is mutually intelligible throughout a majority of the Arab world. The following research will discuss the different spoken Arabic language of Khadijah Salah and Riad Sleiman, specifically the changes in the way the mother says a word versus the way her son says the word, in present day. The focus is on the regional Palestinian dialect of Khadijah Salah and the way that spoken communication transferred to her son Riad Sleiman as he lived primarily in a household of a specific dialect through his first 17 years of life in Kuwait. When discussing the "spoken" Arabic it is important to understand what that encompasses. Spoken Arabic is the different varieties of Arabic that Arabs use for daily communication. There are multiple ways the Arabic language can be broken down to be understood based on the "regional dialects." With regional dialects, it is not only the location of the particular dialect but also sociolinguistic factors that can cause a change in the specific dialect so living in urban areas versus country areas in a particular country can cause a difference in one Arabic dialect. Organizing the groups of regional dialects tends to be hard because while each of these dialects of Arabic can be categorized as "Arabic," some are not mutually intelligible. For example, a Jordanian may easily understand the dialect of a Syrian's spoken Arabic but may have a harder time understanding a Moroccan's Arabic. To categorize the regional dialects of Arabic, we will be using the breakdown suggested by Zaidan and Callison-Burch (2014): Egyptian, Levantine, Gulf, Iraqi, and Maghrebi. This research focuses on the dialects in the areas categorized as the "Levantine" area which would be the countries of Palestine, Jordan, Syria, and Lebanon. While the participants of this research relocated from Palestine to Lebanon to Kuwait (which is located in the Gulf area of Arabic dialects), Khadijah and Riad's dialects are specifically categorized as "Levantine." Levantine Arabic is also familiar throughout the Arab world.

3.1 *Levantine* 'Levantine Arabic' denotes the Arabic spoken by some 40 million people, the natives of Lebanon, Syria, Palestine, Jordan and some parts of Israel inhabited by native Arabic speakers" (McLoughlin, 2009). The Levantine dialect varies within the category itself because it houses multiple dialects like the Palestinian dialect, the Lebanese Dialect, the Syrian Dialect, etc. This will be the main language category focused on throughout this paper.

3.2 *Palestinian Arabic* Palestinian Arabic is spoken mainly in Palestine, but also by the many Palestinian populations who have moved due to the conflict. Palestinian Arabic falls under the Levantine Arabic category. This dialect is closest to Classical Arabic, which is more formally used in books and news reports, and can be understood by many Arabic speakers. Khadijah speaks the Palestinian dialect of Arabic.

3.3 *Lebanese Arabic* Lebanese Arabic falls under the category of Levantine Arabic. It is mostly spoken in Lebanon. It borrows from the Arabic, English, and French languages. Morphologically speaking it is a simple form of Arabic with no mood and case markings. Many people who hear the Lebanese dialect, in comparison to other Arabic dialects, would say that it has a very 'sing-song' quality.

3.4 *Kuwaiti Arabic* Kuwaiti Arabic does not fall under the category of Levantine Arabic; it actually falls under the category of a Gulf Arabic Dialect. It is mainly spoken in Kuwait. Kuwaiti Arabic is known to have borrowed much of its language from other languages due to the vast immigration into Kuwait.

3.5 *Arabic in the United States* According to the U.S. census, 1.5 million people in the U.S. between 2006-2010 were of Arab ancestry. Arabic is also one of the fastest growing languages in the U.S. Many dialects of Arabic can be found throughout the U.S. which can be a result of immigration.

4 Participants in the Study

In this section, we introduce the two main participants in the study- Khadijah Salah and Riad Sleiman.

4.1 *Khadijah Salah* Khadijah Salah, approximately 87 years old[1], is the direct source of a majority of this research. She was born in Palestine where she lived there until she was about 13-15 years old and that is where she picked up her native language of Arabic. Khadijah does not know her exact age because of the lack of documents but she has approximated 87 years old. Khadijah's primary language is Arabic, which she developed in Palestine, making her dialect primarily Palestinian. She is monolingual even though she has lived in America, with Riad Sleiman, since 1992. Khadijah provided her primary accounts of her experience with the Israel Palestine Conflict, and was the primary participant when comparing the data collected, between her data and her son, Riad Sleiman's data.

4.2 *Riad Sleiman* Riad Sleiman, 53 years old, is the biological son of Khadijah Salah and Najib Sleiman. He was born in Baalbeck, Lebanon but grew up for his first 17 years in Khaitan, Kuwait. He spent a majority of the school years in Kuwait, but the family would spend time in Lebanon for the summers. He left his family at the age of 17 to pursue a degree in Aerospace Engineering at Wichita State University. He married Kimberly Faye Hernandez, and brought his parents over shortly after that. He and his wife had 5 children in Wichita, Kansas where he currently resides. Riad is bilingual in both Arabic and English.

5 Dialect Contact

The main aim of this paper is to study the linguistic variation resulting because of language and dialect contact. The adoption of an urban identity through language is very common in the Arabic world (Habib, 2010). Due to the diglossic situation in the Arabic speaking world, some dialects are considered more prestigious than others. Moreover, identity has a huge role to play in making speakers decide which form to use in a particular setting. Thus, an Arabic speaker may adopt an urban form to sound more city-like yet retain their rural form in a different setting, for example at home. This study seeks to answer the following questions:

a) How do social factors such as age, gender, social class, residential area, education, influence the choice of certain speech sounds in Arabic?

b) What kind of variation and change is taking place due to dialect contact between rural and urban speakers?

c) How has displacement due to war influenced language maintenance and language shift?

In particular, we are interested in looking further into aspects of language convergence and assimilation.

5.1 *Convergence* This research explores some aspects to language convergence. This linguistic change plays a role in the results of the comparison between the two spoken forms of Arabic below. Riad was born in Lebanon but grew up in Kuwait. He attended his schooling in Kuwait, spent his summers with cousins in Lebanon and then spent all his years past 17 in America. It is easy to see the interference in his language as different words started resembling the dialects of the countries he was in.

5.2 *Assimilation* This research also explores some avenues of language assimilation in which the younger generations adopt certain features of the dominant language, in this case English. As the father, Riad lived in America and was educated here, he could have adopted some features of English that is missing from the grandmother Khadijah's language. Moreover, Khadijah is monolingual, despite living in America for many years.

[1]Khadijah's birth records were lost in the displacement, and there are no written records of her date of birth or age.

5.3 *Data Collection* The data was collected over 6 months throughout 2017. It is a comparison between the spoken Arabic of Khadjiah Salah, who speaks a primarily Palestinian Dialect, and Riad Sleiman, who speaks a mix of dialects but mostly Palestinian dialect as well, considering that Riad was raised by both parents for 17 years of his life. Also, it was decided to focus on one family specifically to avoid the numerous backgrounds and varieties of language, if more participants from other families were included. These two participants were chosen because they are the most accessible people for the research and for observing them in the everyday environment of their shared home. Everyday utterances were audio-recorded, transcribed, analyzed into a database specifically designed to observe the differences between the two participants. The recordings took place in the family's home in Wichita, Kansas. Some focus areas included the different pronunciations of words, or the changing of words all together.

6 Results

Throughout the research it was found that there were distinct changes in the way the two participants were speaking. While many factors play a role in this, it is important to note that the sole purpose of their movement from Palestine was due to the aforementioned conflict. This displacement from Palestine resulted in this family's movement which further resulted in the family's changing of Arabic through the two individuals, mother and son. Five phonological processes were noted throughout the research: Deletion, Substitution, Breaking of Consonant Cluster, Palatalization, and Voicing.

6.1 *Deletion* Phonological deletion refers to the process that affects the syllable structure due to the omission of one or more sounds. There are different kinds of deletion, such as consonant cluster reduction, initial consonant reduction etc. In the data elicited, we found instances where Riad would delete letters in words that Khadijah would pronounce them. In the two instances identified below the deletion occurred specifically at the beginning of the word, i.e. initial consonant deletion. This is shown in Table 1.

Number	**Khadijah**	**Riad**	**English Translation**
1	**M**enta Shayeff	Enta Shayeff	You see
2	**(Q)**alb	Alb	Heart

Table 1: Data showing deletion of initial consonant sound

In the phrase 'you see' and the word for 'heart', Riad deletes the initial consonant. In the case of 'you see', he deletes the consonant [m], and in the case of 'heart', he deletes the consonant [q], the voiceless uvular stop. This change in the pronunciation of the words can be attributed to the fact that Riad is primarily speaking a different dialect, namely the Kuwaiti dialect. Whereas, Khadijah is speaking the Palestinian dialect. This can be seen in the following sentence from the Kuwaiti dialect, where the word for 'you' starts with the vowel, similar to Riad's utterance.

(1) anti/enti matchufi
you see
'Can't you see?'

However, the Kuwaiti Arabic word for 'heart' is 'galb', where the original Palestinian [q] is voiced. One of the characteristics of Palestinian Arabic is the retention of the pronunciation of the Modern Standard Arabic [q], which is pronounced as /'/ or with the glottal stop [ʔ] in most urban dialects (Jarrar et al., 2017). Thus, in this case, Riad could actually be deleting the initial consonant and using a more urban dialect. Khadijah is speaking the Palestinan dialect, where she pronounces the [q] word initially.

Thus, here we see instances where Riad is speaking a different dialect from the mother's dialect. By virtue of having grown up in Kuwait, Riad's Arabic is influenced by the Kuwaiti pronunciation which is different from the Palestinian pronunciation. We also find evidence that Riad is using a more urban dialect, compared to his mother.

6.2 *Substitution* In this phonological process, one sound is substituted for another sound in a systematic way. In the data collected, there was an instance where we found a change of sound from [w] to [f]. The [w] was pronounced when Khadijah was saying 'wen' which means 'where' in Arabic. Riad pronounced 'where' with a distinct [f] sound when he said, 'fen'.

Khadijah	**Riad**	**English Translation**
Wen rayha al youm	**F**en rayha al youm	Where are you going?

Table 2: Data illustrating substitution of the initial consonant sound

Here, Khadijah again is using the Palestinian word for 'where' which is [we:n], whereas Riad is using the Kuwaiti dialect where the word for 'where' is [fe:n] as shown in example (2).

(2) fen agrab jameya
where nearest shop
'Where is the nearest shop?'

Thus, we find again that Riad's speech has more in common with the Kuwaiti dialect of Arabic, than the Palestinian dialect.

6.3 *Epenthesis* Another instance of phonological deletion is epenthesis. In this process, a consonant cluster is simplified by the insertion of a vowel. This breaks up the consonant cluster and makes it easier for the speaker to pronounce. It was discovered that Riad says the word 'tomato' in a specifically Lebanese dialect, where the word for 'tomato' is 'banadora'. There is an insertion of the vowel [ə] in this word, where Khadijah maintains the consonant cluster [nd] when she says it.

Khadijah	**Riad**	**English Translation**
Ba**nd**ura	Ban**a**dura	Tomato[2]

Table 3: Data illustrating epenthesis of [ə] in medial position.

Epenthesis is obligatory, in Lebanese Arabic, in coda clusters where an obstruent is followed by a sonorant and it is optional in all other coda clusters (Hall, 2011). Thus, again we find evidence that Riad is speaking a different dialect. In this case, he is speaking the Lebanese dialect. Recall that the family would spend summers in Lebanon. Riad could have picked up this feature of the language during one of the family's visits to Lebanon, during Riad's formative years.

Here, we find that although 'bandura' is the Levantine Arabic word for 'tomato', which is prevalent in Khadijah's Arabic, Riad is using the specific Lebanese Arabic pronunciation with the epethetic vowel breaking the consonant cluster. During the Lebanese Civil War (1975-92), the pronunciation of the Levantine word for 'tomato' served as a clue to one's ethnic/national identity in a war situation (Suleiman, 2004). In some cases, the pronunciation served as a boundary setter, a matter of life or death.

6.4 *Palatalization* In the data collection, we found an instance when Riad says the word for the quantifier "a lot of," he inserts the palatal approximant [j] in the middle of the word. In this case, Riad is using the palatal lateral approximant instead of the lateral approximant [l] that Khadijah uses.

Khadijah	**Riad**	**English Translation**
Malaneh	Mal**y**aneh	A lot of

Table 4: Data showing palatalization of the lateral approximant

[2] Note that the word for 'tomato' could be a borrowing from the Italian word for tomato 'pomodoro'. The Modern Standard Arabic word for 'tomato' is 'tamaatim', a derivation of the native American Nahuatl word 'tomatl'.

The word used by Khadijah is a Palestinian word. Whereas, Riad is using either the Lebanese or the Kuwaiti dialect, with the insertion of the palatal [j].

6.5 *Voicing* The phonological process that was most common in the data elicited is voicing. In the words below, the voiceless fricative [Θ] that Khadijah pronounces becomes the voiced alveolar [d] in Riad's dialect.

Number	Khadijah	Riad	English Translation
1	Yokh**th**u	Yokh**d**u	They are taking
2	**Th**ahab	**D**ahab	Gold
3	U**th**un	U**d**un	Ear
4	A**th**an	A**d**an	Call to prayer

Table 5: Data illustrating voicing of the voiceless fricative [Θ] in initial positions

As can be seen, there are different syllabic positions in which the voicing change happens. In examples, (1), (3) and (4), the change happens in the onset position of the second syllable. Whereas in example (2), the voicing happens in the onset position of the first syllable. Here again, Khadijah has preserved the [Θ] in the Palestinian dialect from the original Modern Standard Arabic. Whereas, most urban dialects use the voiceless [t]. Thus, we can hypothesize that here there are actually two phonological processes under play. First, Riad substitutes the voiceless fricative [Θ] with the voiceless alveolar stop [t]. This process is called stopping. Then, he voices it with the voiced alveolar stop [d]. Here again, Riad is using a more urban dialect compared to Khadijah.

7 General Discussion

Overall, we found many phonological differences in Riad's speech when compared to Khadijah's speech. We can conclude that Riad is speaking a more urban dialect, compared to Khadijah. Although, Khadijah has been displaced from Palestine to America, via Lebanon and Kuwait, she is trying to maintain her birth language, the language of her Palestinian identity. She was married to a Palestinian, and therefore even during their stay in Lebanon and Kuwait, her primary dialect for communication was the Palestinian dialect. By the time she relocated to America, she was in her 60s. She had no need to learn English, as the primary language of communication at the family's home in Wichita is Arabic. Thus, she continued to be monolingual and integrated, instead of assimilated into each of her displaced settings.

Riad, on the other hand, was born in Lebanon and was primary educated in Kuwait. Although, he was exposed to the family's Palestinian dialect at home, he interacted with friends and peers who spoke one of the more regional dialects. Moreover, his formal education was in Kuwaiti Arabic. We have seen evidence that social variables such as education, age, gender, has influenced Riad's choice of speech sounds in Arabic. Riad, by virtue of moving to America during his early adult years, has undergone language shift. He now speaks a more urban dialect of Arabic. Notions of convergence and assimilation play a role in the Arabic he speaks.

We further speculate on some of the phonological changes we observed during the data collection. There are many possibilities and predictions that can be made as to why these changes are happening. With the voicing change of the voiceless fricative [Θ] to the voiced alveolar [d], there is the possibility that the way these words are spoken are becoming more of an urban form of speaking, in Riad's dialect. In English, there are instances where, "Who's that?" has become "Who dat?," in modern day urban speech.[3] It is possible that due to Riad's arrival in America at an early age of 17, he adopted a more urban way of speaking English in order to assimilate with the new, English speaking environment.[4] The insertion of the palatal approximant [j] in his speaking of the word "malyaneh"

[3] Francis X Connor (p.c).

[4] Note that Riad Sleiman spoke very little English when he moved to Wichita. He had some English as part of his high school language instruction but he came to America with minimal English speaking abilities.

versus Khadijah's "malaneh" can also be argued to occur because of urbanization. "Malaneh" is the more colloquial form, used in Palestine Arabic, whereas "malyaneh" might be the more urban form. His influence of learning English and the English way of speaking may have changed the way he pronounces things in Arabic as well. Khadijah likely clings to her Palestinian dialect because of her continuation of speaking the Palestinian dialect with her husband and children until the husband passed away in 2002. Also, it is likely in social settings, such as with friends or family, she was also able to continue to use her Palestinian dialect with no social repercussions. She also remained in Arabic speaking countries where she did not need to adapt to a new language so she would be able to continue her speaking of Arabic without any interruptions.

In our study, we found that age, gender, social class, education, influence the choice of certain speech sounds in Arabic. Riad's age of entry into America, his gender, and his educational backgrounds have led him to adopt an urban form of Arabic. Riad being the younger speaker is more inclined to use urban forms such as the [ʔ] instead of the older form using [q]. Thus, there is a linguistic shift towards the more prestigious, urban form. The older speaker, Khadijah, shows language maintenance of their native form [q]. This is an indication of pride in her rural identity and solidarity with her rural roots. Thus, we have seen that war and displacement has resulted in the younger participants shift in language use, whereas the older participants maintain their identity and language.

8 Further Research

These discoveries have opened up many avenues to further expand this research. In expanding this research, there are some central areas to focus on that can also be explored as factors for this research. Expanding the participant pool for data is a primary goal for future research in order to explore the trends on a larger scale with a variety of participants. There are many Palestinian refugees around the world especially in America. Later generations of Palestinians have been born in America, and they primarily speak English, so it is important to look at the different ways these people are speaking the Palestinian dialect/Arabic, or if they are even speaking Arabic at all. Many of the people who have lived through the 1948 conflict are aging quickly so it is important this research and data is being collected while they are alive. Another factor to include in further research would be generational and gender characteristics that play a role in the development of the language. These two factors could play a role in expanding this research and something to look at specifically in future research This research also focused on a Muslim family. Religion could be a social variable to investigate, as many Christian Arabic communities speak a different dialect. Lastly, regional areas throughout Palestine differ in dialect which could be another factor to look at when exploring the changes through the generations.

The authors plan to expand the study by looking at other Palestinian families in Wichita. There are roughly 300 Palestinians in Kansas, with Sedgwick County housing the largest amount of them, according to the Arab American Institute Foundation's census taken between 2005-2009. We plan to include more participants in each social variable, by providing them with a questionnaire.

9 Conclusion

The development of the Arabic spoken in this family is still changing till this day. With the large influences on the terminology they currently possess, such as their displacement to different countries, it is no surprise that the Arabic this family is currently speaking is very unique. The overall uniqueness of the Arabic spoken in this family can be traced back to each of the lands in which the grandmother sought refuge in order to keep her and her family safe and together.

The results of this study could aid English language learners whose Arabic is their first language. Professional development tools should reflect subtle variations, depending on different social variables, as this will aid language teachers understand different customs for helping students transition to English, thereby enriching the second language learning pedagogy.

References

García, Justin D. 2014. "Arab-Israeli Conflict." *Salem Press Encyclopedia*, January. EBSCO *host.*

Habib, Rania. 2010. Rural Migration and Language Variation in Hims, Syria. *SKY Journal of Linguistics* 23 (2010), 61-99.

Hall, Nancy. 2011. Vowel Epenthesis. In, Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.) *The Blackwell companion to phonology*. Malden, MA & Oxford: Wiley-Blackwell. pp.1576--1596.

Jarrar, Mustafa, Habash, Nizar, Alrimawi, Faeq, Akra, Diyam, and Zalmout, Nasser. 2017. Curras: an annotated corpus for the Palestinian Arabic dialect. *Language Resources and Evaluation*, 51 (3): 745-775.

McLoughlin, L. J. 2009. *Colloquial Arabic (Levantine)*. Routledge.

Morris, Benny. 2004. *The Birth of the Palestinian Refugee Problem Revisited*. Cambridge: Cambridge University Press.

Pappe, Ilan. 2006. *The Making of the Arab-Israeli Conflict*, 1947-1951. I.B. Tauris & Co Ltd.

Suleiman, Yasir. 2004. *A War of Words: Language and Conflict in the Middle East*. Cambridge: Cambridge University Press.

Teveth, Shabtai. 1990. The Palestine Arab refugee problem and its origins. *Middle Eastern Studies*. 26 (2). Pp. 214-249.

Zaidan, Omar.F. & Callison-Burch, Chris. 2014. Arabic dialect identification. *Computational Linguistics*. 40 (1). Pp. 171-202.

Obviation in Karuk

Ashley Tomlinson
Boise State University

1. Introduction to the Karuk People and their Language

Similar to other languages, Karuk uses person hierarchy and inverse marking to indicate subjects and objects in transitive actions. These features are used to indicate topicality among participants in a given action. In addition to these features, Karuk also uses obviation. Obviation is also seen in Algonquian languages which sets the expectations for how obviation works. However, there are a few unique differences in the function of obviation in Karuk compared to its function in Algonquian languages. The function and use of the obviative marker, *'îin,* in Karuk will explored in this paper. Before exploring this marker, the Karuk language and the aspects of person hierarchy and inverse marking will first be discussed as well as the use of obviation in Algonquian languages.

1.1 *The People.* Karuk is the language spoken by members of the Karuk tribe in northern California along the Klamath river. Karuk itself means 'upriver' representing the upstream location these people occupy along the river. This area is between the Seiad valley to the north and Buffalo Creek to the south. Traditionally, the Karuk people lived in small villages without chiefs. In their villages, the people had housing structures divided for the women and men. The women and children occupied what was referred to as the living houses while the men occupied the sweathouses.

As the Karuk people were known as the 'upriver people,' fishing was a large part of their lives and subsistence. They primarily fished salmon, which became a large staple in their diet. They primarily fished using the dip-netting method, which has been passed down through the generations. In addition to fishing, the Karuk people foraged in the nearby forests for acorns and other plants in addition to hunting other game. Tobacco-growing was common among the Karuk people and is still used in ceremonial events. Ceremonies and ceremonial taboos were a large part of their lives. Other common activities seen among the Karuk people were building fishing platforms and crafting basketry (Bright).

1.2 *The Language.* Karuk is a language isolate. There were an estimated 1500 speakers prior to European contact. In 1957, William Bright estimated that there were about 100 speakers. This dropped to 15 remaining fluent speakers in 1989, and today there are less than ten fluent speakers left. Similar to many other Native Americans, the Karuk children were sent to schools and punished for the use of Karuk and taught English, greatly reducing the number of children learning Karuk until it was no longer taught to their children at all (Bright). However, there are strong revitalization efforts for Karuk today and a cohort of second language speakers as a result.

Karuk is a polysynthetic language containing most of its morphology in the verb, which consists of a complex person-marking system. The verbs also carry a large system of direction suffixes. Rather than carrying terms for cardinal direction as in English, Karuk's directional suffixes refer to motion in relation to a point of reference. Examples given by the Survey of California and Other Indian Languages (WALS) is, "uphill away from that point of reference," and "horizontally away from the center of a body of water." The nouns in Karuk, however, lack case and number. An interesting feature of Karuk that will be investigated here is its hierarchical person system and accompanying inverse and system of obviative marking seen throughout the language. These features are noted in a relatively small number of languages, but play a large role in the sentence structure of Karuk for subject-object interactions.

* The author thanks Dr. Tim Thornes for guidance and support throughout this research and the audience at WECOL 2017 for helpful feedback.

Proceedings of *WECOL 2017*

2. Grammatical Structure of Subject/ Object Interactions

2.1 *Person hierarchy in Karuk* Person hierarchies in language are typically found to rank first person over second person and second over third person (1 > 2 > 3). However, in Karuk a slightly different pattern is followed with Karuk person hierarchy being second person plural > first person > second person singular > third person (Macaulay 2000). Karuk uses pronominal affixes, generally a prefix, to mark the higher ranking participant in a sentence. This hierarchy is used to determine which participant is referenced in a given interaction. For example, "since 1 outranks 3, we expect the first-person participant to be the one who is referenced in these cases---that is, we expect the object to be referenced when a third person acts upon a first, and the subject to be referenced when a first person acts upon a third" (Macaulay 1992).

Although the participant of the higher ranking is the one referenced in the interaction regardless of their role as subject or object, the pronominal affix will vary depending on which role the higher ranked participant takes. The subject-object combination of an interaction in relation to this hierarchy will determine which pronominal affix is used. To represent the different combinations of subject-object interaction, the pattern X > Y will be used showing X as the subject that is acting on Y the object (Macaulay 1992). For example, 1sg > 3sg shows first person singular acting on third person singular such as in the sentence, "I told her."

According to Macaulay (1992), there is another factor besides the hierarchy and subject-object combination that also determines which pronominal affix is used. Whether the interaction is positive (P), negative (N), or optative (O) in mood will affect what prefix is used as well. The individual prefixes that are used in Karuk can be seen in Table 1:

Subject/Object	Positive Mood	Negative Mood	Optative Mood
1sg > 2sg	*nu'-*	*kin-*	*nu-*
1sg > 3sg	*ni-*	*na-*	*kan-*
1sg > 3pl	*ni-*	*na-*	
1pl > 2sg	*nu'-*	*kin-*	*nu-*
1pl > 3sg	*nu'-*	*kin-*	*nu-*
1pl > 3pl	*nu'-*	*kin-*	*nu-*
2sg > 1sg	*na-*	*na-*	*na-*
2sg > 1pl	*kin'-*		*kin'-*
2sg > 3sg	*?i-*	*0-*	*0-*
2sg > 3pl	*?i-*	*0-*	*0-*
2pl > 1sg	*kana-*		*kana-*
2pl > 1pl	*kin'-*		*kin'-*
2pl > 3sg	*ku-*		*ki-k-*
2pl > 3pl	*ku-*		*ki-k-*
3sg > 1sg	*na-*	*na-*	*na-*
3sg > 1pl	*kin'-*		*kin'-*
3sg > 3sg	*?u-*	*0-*	*kam-*
3sg > 3pl	*?u-*		*kam-*
3pl > 1sg	*kana-*		*kana-*
3pl > 1pl	*kin'-*		*kin'-*
3pl > 3sg	*kun-*		*kun-*
3pl > 3pl	*kin'-*		*kin'-*

Table 1. Adapted from Macaulay 1992

2.2 *Inverse marking with the suffix –ap.* When an action or event in a sentence follows the hierarchical order of persons it is considered direct. An example of this is the sentence "I told her". The first person is the higher ranked participant in this interaction and they are acting on the third person who is lower ranked, directly following the person hierarchy. When the interaction does not follow this pattern it is inverse (e.g. "She told me" 3sg > 1sg). According to Macaulay (1992), "inverse markers are morphemes

which reverse the expected subject-object pairing for a transitive verb." Karuk marks this inversion with the suffix *-ap*. From Macaulay's observations, -ap converts a first- or second-person subject into an object. This suffix is typically used with the pronominal prefixes. When the prefix *?i-* is used a second-person singular subject acting on a third-person object is marked. However, when the suffix -ap is added it reverses the action to a third-person subject acting on a second-person singular object (Macaulay 1992).

Here are a few examples of Karuk sentences using these affixes from Macaulay (1992). Consider:

(1) *vura puna?ave-sara*
EMPH NEG=1SG/3PL(NEG)-eat-FUT-NEG
'I won't eat them'

In this sentence, the prefix *na-* is seen to represent the first-person singular (I) who is acting on (eat) the third-person plural (them) in a negative series (won't). This action follows the Karuk person hierarchy and so the suffix *-ap* is not seen. Observe the following, by way of comparison:

(2) *ta' p-u- kik-ta'pkump* ***-ap***
PERF NEG=2PL/3PL(NEG)-like-INV
'They do not like you.'

In this sentence the -ap suffix is used to indicate that the person hierarchy is violated with the third person plural (they) acting on the second-person plural (you) in a negative series (do not).

In both of these examples the addition or removal of *-ap* would change the interaction. If the first example added the *-ap* suffix, the meaning of the sentence would become 'they won't eat you.' If the second example dropped the suffix the meaning would shift to 'you do not like them.' As Karuk is a polysynthetic language and does not specify a specific word order pattern to follow, inverse markers along with the pronominal prefixes become important identifiers. Additionally, obviation also becomes an important identifier between proximate and obviate participants when both subject and object are third person. This feature is used in Karuk, but prior to exploring how it is used in Karuk obviation in Algonquian languages will first be discussed.

2.3 *Obviation in the Algonquian languages.* Obviation was first documented and discussed among Algonquian languages. Obviation is used as an obligatory marker for third person characters in a sentence. This is done to distinguish which third-person character is proximate or more prominent from all other third-person characters who are considered obviate or less topical (Aissen 1997). Essentially, the obviate is used to further break down the person hierarchy by distinguishing the different level of importance among animate third-person characters who otherwise occupy the same place on the hierarchy scale (Macaulay 2000).

The participant in the subject role takes the higher ranking in the person hierarchy compared to the participant in the object role, when both participants are third person. In these cases, the action is considered direct and no marker is used. However, when the proximate participant takes on the object role, the action is considered inverse. When this occurs, the obviate is employed to indicate the indirect form and to highlight the agent as obviate rather than proximate.

In Algonquian languages, obviation is used when the action is in inverse form with the obviative aligning with the subject and the agent role (Aissen 1997). According to Aissen, the use of obviation will vary on the languages preferred subject choice. In Algonquian languages, such as Fox and Plains Cree the preferred subject is the participant with the higher semantic role. (Ibid). This is the agent over the patient. In cases where the patient is the proximate and more topical participant, a passive structure is not used as the agent remains the preferred subject. In order to identify the inverse alignment and recognize the agent as less topical, obviation is used (Ibid).

Obviation in Ojibwa, another Algonquian language, is used when possession is used between people as in the first example and when both participants in a sentence are third-person by adding the obviative marker to the verb as well as the obviate character. The obviate marker is bolded in the examples below:

(3) *gimaa w-gwis-**an***
chief 3-son-OBV
‘The chief’s son’
(Mithun 1999)

(4) *aw aniniw w-gï-wâb-am-à-**n** niw kwêw-**an***
DEM.PROX man 3A-PST-see-TA-DIR-OBV DEM.OBV woman-OBV
‘The man (prox.) saw the woman (obv).’

(5) *aw kwëw w-gi-wäb-am-igw-**an** niw aniniw-**an***
DEM.PROX woman 3A-PST-see-TA-INV-OBV DEM.OBV man-OBV
‘The man (obv.) saw the woman (prox.)’
(Zúñiga 2014)

2.4 *Obviative Marking with ‘îin in Karuk.* In addition to the inverse marking system in Karuk that marks when an interaction violates the person hierarchy, the language also has a system that marks a non-main character in the interaction. Obviation is marked with *‘îin* in Karuk and only occurs with transitive verbs when something is done directly to someone or something. *‘îin* was referred to as an agentive position by William Bright who has done the bulk of the primary linguistic research on the Karuk language. He further defines it as, “. . . marking the [noun phrase] subject of a transitive verb with animate subject and object” (Macaulay 2000).

Throughout examples of *‘îin* found in Karuk, it serves as a marker that “. . . tags a subject [noun phrase] as a peripheral character acting on the main character of the narrative” (Macaulay 2000) and appears to follow an obviative pattern. Obviation is a reference tracking system that identifies the central third person character as central (what is often referred to as "proximate") while all others are considered obviative. In a sense it is taking the person hierarchy a step further by placing a ranking amongst third-person participants.; identifying a “fourth-person” character (Macaulay 2000).

Specific conditions are required for the use of *‘îin* to occur.:

- The verb must be transitive,
- The subject and object must be animate,
- The subject is virtually always third-person,
- A full [noun phrase] subject is present,
- A full [noun phrase] object is usually not present, and
- The object must be the main character in the narrative (Macaulay 2000).

Unlike other languages that use an obviative marking system, Karuk has three distinct differences with its obviative system. The first is that all *‘îin*-marked sentences in Karuk must contain a proximate argument while other languages use obviation for either a proximate/obviative argument or two obviative arguments. The main use of obviation is to distinguish between a third-person main character and another third-person participant who is not a main character. This brings up the second difference in Karuk. While a majority of Karuk sentences with an *‘îin* marking contain two third persons, there are some that use *‘îin* to mark the subject even though the object is a first- or second-person. Obviative marking is not expected in these situations as the object is not a third-person and, according to the person hierarchy and pronominal affixes, the reader/listener already knows that the first- or second-person outranks the third-person subject. With this difference, it appears that obviative marking in Karuk has a slightly adapted purpose for the marker—it reminds the listener that the main character is being acted on and that the subject is not the most important participant of the sentence. Finally, the obviative marker in Karuk, *‘îin,* is a postposition rather than a suffix that is on the noun or verb (Macaulay 2000).

Some of these differences can be seen in the following examples of Karuk sentences with obviative marking:

(6) *ishyuux kun-iyvur-kurih-vu-tih pa-chishiih **‘îin***
elk 3PL/3SG.POS-chase-into.water-to-DUR ART-dog SUBJ
‘The dogs used to herd elk into ravines.’ (Macaulay 2000)

This sentence follows the general structure of obviation. There are two third-person participants, 'the dogs' and 'elk,' with a transitive verb, 'herd.' *'îin* follows behind 'the dogs' as a postposition identifying it as the agent in the sentence and also marking it as the obviative character while 'elk' is the proximate character.

Consider next, the following example with a second person object:

(7) *?i-m ?o-k ke-misa* ***' îin*** *?i-?av-avis-ap*
Outside here monster SUBJ 2SG/3SG.POS-eat-FUT-INV
'The monster outside here is going to eat you.'
(Macaulay 1992)

This sentence represents the unique feature of Karuk to use *'îin* with a second-person object. 'you.' The obviative marker is again seen as a postposition behind the peripheral third-person character agent, *ke-misa* 'monster.' The second-person 'you' ranks higher than the third-person 'monster' and is the main character in the action. Different from sentences containing two third-person participants, sentences that use obviation with a first- or second-person object also contain the pronominal affixes with the personal prefix and suffix *–ap*. This is seen in this example. *?i?ave-sap* represents 2sg > 3sg (positive)-eat-FUT-INV. The personal prefix *?i-* is added to show the second singular person acts upon the third-person singular while the suffix *–ap* is added to inform the listener that the action taking place is inverse with the third-person acting on the second-person.

As seen in both examples above, both the participants in the sentences are animate. This is one of the conditions listed earlier about the use of *'îin* in Karuk—both participants have to be animate such as humans and animals. The obviative marker cannot be used if one of the participants is inanimate such as a tool. For example, in a sentence such as, "the blade nicked him," no obviation would be used even though a non-main character is doing the action and the object outranks the subject because the blade is inanimate. Another trait that is seen in sentences with *'îin* is that they are all inverse. The person hierarchy discussed earlier is the structure that obviative sentences follow. Since both participants are, generally, third-person the marker is needed to show the inverse action that is taking place by marking the non-main character doing the action.

As mentioned earlier, the use of *'îin* with non-third-person participants is unique in Karuk compared to the use of obviation in Algonquian languages. This difference appears to be explained by looking at narratives. In the creation story, 'A Trip to Indian Heaven,' the main characters throughout the story are a woman and her lover. The first appearance of *'îin* in the story is after the man encounters a group of people on his journey to save his sweetheart. The group is on their way to see the man's lover in Indian Heaven and the man asks them to tell her that he is there. When this group approaches the woman and one of them speak to her, *'îin* is seen marking them as obviate and the woman as proximate:

(8) *xas yitha* ***'îin*** *kunchuphuunish*
and.then one OBV he.talked.to.her
'Then one of them spoke to her.' (Lang 1994)

Prior to this point, the obviative marker has not been used and the woman has not recently taken part in any interactions.

In this narrative, the woman is the main character and when one of the member's in the group speaks to her, as in (8), *'îin* appears to mark them as less topical than the woman. After this interaction, the narrative shifts to focus back on the woman and her lover as they journey back to their home. Later the woman and man have a child and *'îin* appears when they each first speak to the child. Each parent is marked obviate in relation to their child which can be seen the examples below:

(9) *xas pa-mu-'akah* ***'îin*** *tah kun-ipeer*
and.then ART-his-father OBV PERF 3PL/3SG.POS-tell
'his father told him'

(10) *xas mu-taat* ***'îin*** *tah kun-ipeer*
And.then his-mother OBV PERF 3PL/3SG.POS-tell
'Then his mother asked him'
(Macaulay 2000)

At this point, the boy appears to take on a more prominent role as his actions later affect the outcome of the story. Although, there are multiple interactions between two third-person participants throughout this narrative, *'îin* only appears four times. Each time it appears, there is a shift between the character in the agent role and the character who is more prominent. These examples appear to point to *'îin* as functioning pragmatically to mark when the proximate character changes, regardless of whether they are first-, second-, or third-person.

This can also be seen in the narrative, "The World Renewal Ceremony at Katimin." In this narrative, the main character is 'we' or a first-person plural participant. Throughout the narrative, the main character is the only participant referenced. Part way through, a third-person plural participant is introduced, 'everyone.' The clause that introduces them is:

(11) *koovura-**'îin** kin- immustih*
All OBV 3PL/1PL.POS-look.at
'Everyone looked at us.'
Taken from the narrative "The World Renewal Ceremony at Katimin" (Macaulay 2000)

In this clause, the third-person plural participant, 'everyone,' is taking on the agent role and acting on the main character, the first-person plural participant, 'we.' The obviate marker, *'îin,* appears marking the new participant as less topical than the participant in the object role. The new participant is the subject of the subsequent clauses. All of these clauses are intransitive without any action taking place in regards to the main participant, however, the listener knows that the subject is still less relevant to the story since *'îin* was used to mark as peripheral.

3. Summary and Conclusion

Person hierarchy and participant status are important features in Karuk with the use of pronominal affixes, inverse marking, and obviation. Although a general rule set is seen in examples of inverse and obviative marking in Karuk, there are also examples of exceptions and irregularities that are not fully understood yet. Some examples seen in the language appear to reflect redundancy with the use of an inverse marker as well as the obviative when the object is either first- or second-person. These instances are interesting as the assumption is for obviation to occur when both participants are third-person. If the pronominal affix and inverse marker, *-ap*, both show that the action does not follow the expectations set in the person hierarchy, why would the addition of *'îin* be needed? One hypothesis as to why this irregularity of inverse and obviative marking occurs is that the main function of *'îin* is to signal a change in the proximate character and to remind the listener who the main character is. These features are challenging to study as Karuk is an language isolate, and the number of native speakers continues to decrease. However, the Karuk tribe is currently implementing a strong revitalization program. Developing a clearer understanding of obviative and inverse marking may help to inform revitalization efforts. The function of *'îin* as a discourse marker in establishing shifts between proximate characters appears to explain the redundancy seen between pronominal affixes and inverse marking in transitive clauses. However, the development of this construction still needs to be explored. The continued exploration of how and when *'îin* is used will likely prove to be important with the development of new written narratives, particularly with the indication of topicality of participants throughout the narrative.

References

Aissen, J. (1997). On the Syntax of Obviation. *Language, 73*(4), 705-750. doi:10.2307/417324

Anderton, A. (2006). William Oliver Bright. *Journal of California and Great Basin Anthropology, 26*(2), 117-122. Retrieved from http://www.jstor.org/stable/27825833

Bright, William. "Karuk." (n.d.): 180-87. Web. 6 Feb. 2017.

Bright, W. (1957). The Karok Language. UCPL 13. Berkely: University of California Press.

Gehr, S (2017). *Karuk.* Boise State University. Interview. 4 April 2017.

Hinton, L., & Hale, K. (Eds.). (2001). *The green book of language revitalization in practice*.

"Karuk." *Karuk – Survey of California and Other Indian Languages*. Berkeley Linguistics, n.d. Web. 06 Feb. 2017.

Karuk Tribe (n.d.). Language Program. Retrieved April 06, 2017, from http://www.karuk.us/index.php/departments/language-program

Lang, J. (1994). *Ararapíkva: Creation stories of the people : traditional Karuk Indian literature from northwestern California*. Berkeley, Calif: Heyday Books.

Macaulay, M. (1992). Inverse Marking in Karuk: The Function of the Suffix -ap. *International Journal of American Linguistics, 58*(2), 182-201. Retrieved from http://www.jstor.org/stable/3519755

Macaulay, M. (1993). Reduplication and the Structure of the Karuk Verb Stem. *International Journal of American Linguistics, 59*(1), 64-81. Retrieved from http://www.jstor.org.libproxy.bosestate.edu/stable/1265470

Macaulay, M. (2000). Obviative Marking in Ergative Contexts: The Case of Karuk 'ïin. *International Journal of American Linguistics, 66*(4), 464-498. Retrieved from http://www.jstor.org/stable/1265846

Mithun, M. (1999). The Languages of Native North America. Cambridge: Cambridge Univ. Press. doi:https://blackboard.boisestate.edu/bbcswebdav/pid-4471824-dt-content-rid-22347635_1/courses/1173-14811LING397001/Mithun%20%281999c%29%20Chapter%206.pdf

Pereltsvaig, A. (2011). Revitalization of the Hupa language. Retrieved April 07, 2017, from http://www.languagesoftheworld.info/geolinguistics/revitalization-of-the-hupa-language.html

Romney, L. (2013). Revival of nearly extinct Yurok language is a success story. Retrieved April 07, 2017, from http://articles.latimes.com/2013/feb/06/local/la-me-yurok-language-20130207

Silver, S., & Miller, W. R. (1997). American Indian Languages: Cultural ans Social Contexts . Tucson: The University of Arizona Press.

University of California Publications in Linguistics. http://content.ucpress.edu/chapters/11619.ch01.pdf

"WALS Online - Language Karok." *WALS Online - Language Karok*. N.p., n.d. Web. 06 Feb. 2017.

Zúñiga, F. (2014). Inversion, Obviation, and Animacy in Native Languages of the Americas: Elements for a Cross-linguistic Survey. *Anthropological Linguistics, 56*(3/4), 334-355. Retrieved from http://www.jstor.org/stable/24582253