HOLISTIC LEXICAL STORAGE: COARTICULATORY EVIDENCE FROM CHILD SPEECH

Margaret Cychosz[1]

[1]Department of Linguistics, University of California, Berkeley
mcychosz@berkeley.edu

ABSTRACT

Adult speakers readily decompose words into their component parts. Overgeneralizations in children's early speech (e.g. *goed*) demonstrate that they must share in this ability. However, acoustic evidence from child speech suggests that children do not always break words down, instead storing language in more holistic chunks such as syllables or even entire words.

How are morphologically-complex forms represented throughout childhood? To answer this, we measured coarticulation between the same biphone sequence, [ap], in two environments: 1) within morphemes and 2) across morpheme boundaries in adult and child (age 5-10) South Bolivian Quechua speech. Adult speakers co-articulated less across morpheme boundaries than within root morphemes. This is further evidence that adult speakers decompose complex words. Children, however, coarticulated equally across and within morphemes. This suggests that the child speakers store inflected words more holistically than adults, even in this highly agglutinating language.

Keywords: coarticulation, acoustics, acquisition, field phonetics, morphology

1. INTRODUCTION

Adult speakers can compose novel words efficiently, freely converting nouns into verbs (*monetize*) and verbs into adjectives (*twinkly*). Since the Wug Test [7] , we have acknowledged that young children must share this morphological productivity with adults; if not, they would not be able to extend morpho-phonological patterns to novel word environments.

Despite these assumptions, it is likewise apparent that morphological decomposition, or the deconstruction of morphologically-complex words into morphemes, varies by factors such as word frequency. For example, in adults, highly frequent complex words are less likely to be decomposed [2]. Derived words that are more frequent than the corresponding base form (*disentangle* vs. *entangle*) are also less prone to decomposition [14] (see [18] for alternative explanation).

Furthermore, phonetic production data have called into question the nature of children's early lexical representations. A consistent, though not universal finding is that children coarticulate between adjacent and near-adjacent phones more than adults do [13], [17], [27], cf. [5]. The age of children studied varies (4;0-9;9,) [year;months]. Still, the finding that children coarticulate more than adults suggests that children organize speech more holistically, in syllables or words. Perhaps, for extremely frequent collocations, speech could even be represented in chunks that transcend word boundaries [1]. Anticipatory coarticulation between adjacent phones then decreases as children age; this may represent the individuation and abstraction of language units.

One method of studying the nature of lexical storage is to examine the effects of morphological structure on speech production [3], [10], [18]. A morpho-phonetic relationship is apparent as English /l/ is darker in stem-final position (*coolest*) than affix-initial (*coupless*) [16] and when hetero-morphemic words are temporally longer than identical monomorphemic words (*sighed* versus *side*) [24]. English speaking children show morpho-phonetic interaction as young as 2;0 when morphemic /z/ (*toes*) is longer than non-morphemic /z/ (*nose*) [22] (see also [23]). Consequently, acoustic cues provide implicit evidence of morphological decomposition in adult, and some evidence suggests, child speech.

Here we bring these two lines of research – morpho-phonetic interplay and child coarticulation – together and employ spectral measurements to examine morphological (de)composition in adult and child speech. We measure this in South Bolivian Quechua (SBQ), a highly agglutinating language with over 200 nominal and verbal suffixes. Coarticulation was measured between the adjacent phones [a] and [p] carried in words in one of two contexts: within a root morpheme (e.g. p**ap**a 'potato') or at a morpheme boundary (e.g. llam**a-p**i 'llama-LOC'[i]).

SBQ provides unique insight into interactions of morphology and phonetics. Adult SBQ speakers have a highly flexible inflectional and derivational lexicon: suffixes and roots are abstracted away from the original lexical contexts and are easily rearranged for novel stem+suffix pairings. This process is similar to how speakers of more analytic languages, such as English, arrange novel noun-adjective pairings.

Child coarticulation results suggest that word representations initially develop more holistically. If this result is applied to SBQ, we can predict that child speakers would not decompose complex words. Children would then coarticulate more than adults at morpheme boundaries (e.g. llam**a-p**i 'llama-LOC'), indicating holistic lexical storage. Adults would coarticulate less across boundaries (e.g. llam**a-p**i 'llama-LOC') than within morphemes (e.g. p**ap**a 'potato'), suggesting that they have learned to parse speech segmentally and store it morpho-phonemically. We can tell that this speech pattern comes from experience because adults' phonetic realization of the *exact same* [ap] sequence differs within a morpheme versus across a boundary.

2. METHODS

2.1. Participants

30 children (15 girls, 15 boys) and 10 adults (10 female), all bilingual Spanish-SBQ speakers from a mid-size town in Bolivia, completed a picture-prompted word elicitation task. Child participants were grouped by age: five 5- to 6-year-olds (mean=6;6 [years;months]), nine 7-year-olds (7;7), six 8-year-olds (8;5), five 9-year-olds (9;8), and five 10-year-olds (10;7).[ii] Study hypotheses and design, including the partitioning of age groups, were preregistered in Open Science Framework on September 11, 2018. Registration, raw data, and code to replicate analyses are available in the project.

2.2. Stimuli

To elicit the words, participants were presented with photos of 32 culturally-appropriate nouns that children in these communities recognize (e.g. *house*, *flower*, *cow*). Only words with adjacent [a] and [p] phones in stressed position were analyzed (N=11) (' marks stress; ' marks ejectives) (Table 1). The sequence [ap] was chosen because it occurs within and across morpheme boundaries in many common SBQ nouns that children would know.

Table 1: Words elicited and morphological environment: between or within morpheme.

WORD	TRANSLATION	ENVIRONMENT
'api[iii]	'corn/citrus drink'	within
'papa	'potato'	within
imi'lla-pi	'girl-LOC'	between
juk'u'cha-pi	'mouse-LOC'	between
lla'ma-pi	'llama-LOC'	between
sun'kha-pi	'beard-LOC'	between
t'i'ka-pi	'flower-LOC'	between
uhu't'a-pi	'sandal-LOC'	between
wa'ka-pi	'cow'-LOC'	between
wall'pa-pi	'chicken-LOC'	between
wa'wa-pi	'baby-LOC'	between

Only two within-morpheme tokens could be reliably elicited from the children because we had to ensure that children knew the words. There is no equivalent to the *Macarthur Bates Communicative Development Inventory*, which reports stages of age-normed vocabulary development, for any Quechuan language or Bolivian Spanish. Nor is there a large, naturalistic child-directed speech corpus to infer vocabulary development. We confirmed children's knowledge of the test items in two ways: first, a pre-test confirmed that children as young as 3;0 could name all items. Second, unlike other child elicitation techniques, children here did not mimic a model speaker but instead produced the word spontaneously. If a child did not know a word, it was skipped, though this was infrequent. In this way, we are confident that children knew the words they produced.

2.3. Data collection and analysis

Word lists were recorded with a portable Zoom H1 Handy Recorder at a 44.1 kHz sampling rate. Children in these communities have limited exposure to technology, so instead of eliciting items on a screen, a photo of each item was pasted onto a single page in a binder. For this reason, the words could not be randomized and were presented in the same order for all participants.

Participants first produced all words in Spanish and then repeated the task in SBQ; only SBQ results are reported here. Each of the eleven target words was elicited twice, via two distinct pictures on separate trials. Ideally, this combination would result in 4 tokens**X**11 words. However, due to noise interference from wind and livestock, as well as children's occasional nervousness, some children only produced the target word one or two times. We controlled for this between-subject variability by testing the parameter **Utterance number** in the statistical models; it did not improve model fits.

For each of the items, participants named the item in the photo in a carrier phrase by twice repeating, "I say in the __ two times" (*Noqa nini __-pi iskay kutita*). Most children under 8;0 could not remember the carrier phrase and instead first identified the bare noun (e.g. *llama*). Then, the researcher placed a large plastic toy insect on top of the photo and prompted the child, "Where is the bug?" to which the child produced the word with the correct suffixal carrier e.g. *llama-pi* (*llama*-LOC, "on the llama") two times. The task took 20-30 minutes. All participants were monetarily compensated. Children could additionally choose an item from a toy bag.

Productions were manually segmented in Praat [8]. Much of the child coarticulation literature employs coarticulation measures such as center of gravity or formant transitions. However, measurements of child formants are notoriously difficult to obtain reliably. We closely follow the methodologies employed in [11], who likewise studied child stop-vowel coarticulation, and compute coarticulation via an automatically-extracted measure of spectral change: the difference between averaged

Mel-frequency log magnitude spectra of the adjacent phones. The acoustic signal was first downsampled to 12 kHz. Then, each phone was segmented into 25.6ms frames, with a 10ms step. The Mel-frequency spectral vectors from each phone were then averaged with the resulting vector scaled by the duration of the carrier word to control for speaking rate. Finally, we measured the Euclidean distance between the averaged Mel spectral vector for each [a] and the following [p] for each word:

$$(1) \qquad d_{ap} = \sqrt{\Sigma \left(\frac{\bar{x}_a}{D_{word}} - \frac{\bar{x}_p}{D_{word}} \right)^2}$$

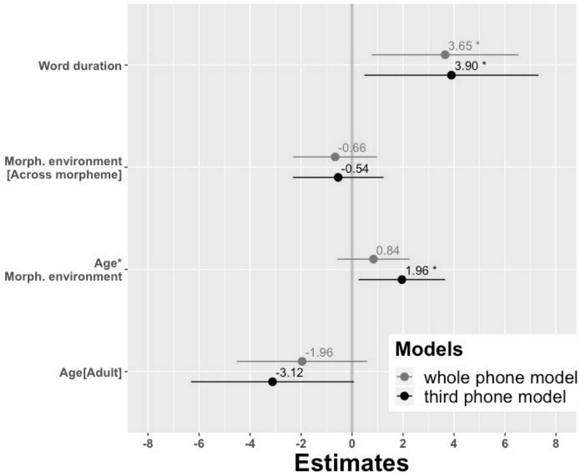
where d_{ap} is the Euclidean distance, $\bar{x}_a$ and $\bar{x}_p$ are the averaged Mel spectral vectors for [a] and [p] from a given word, and D_{word} is the duration of the carrier word.

3. RESULTS

A linear mixed effects regression model was fit to predict the Euclidean distance between [a] and [p] using the lme4 [6] and lmerTest [15] packages in R [19]. Potential model parameters were evaluated using a combination of between-model log-likelihood comparisons, AIC estimations, and *p*-values. Continuous variables were mean-centered. The model included random intercepts for **Word** and **Participant**. Evaluated parameters are listed in Figure 1 under 'whole phone model.'

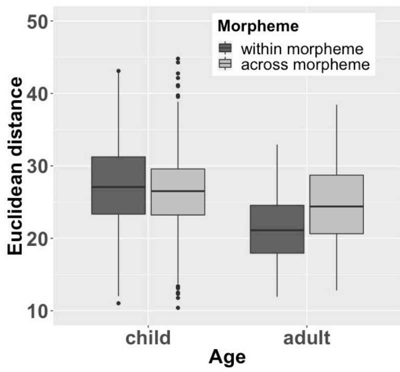
The tested predictors for the model were **Word duration**, **Morphological environment** [across morpheme vs. within morpheme], **Age** [child vs. adult], and the interaction of the latter two. Of these predictors, only **Word duration** improved model fit (β=3.65, t=2.49, p=.013, 97.5% CI=0.34, 6.58). Here a positive beta coefficient indicates less coarticulation between phones in temporally longer words. (The coarticulation metric was also scaled by word duration during measurement, as shown in example [1], but including it in the model maximally controls for the prevalent speaking rate differences between adults and children.) The lack of effect for age or morphological environment indicates that neither children nor adults distinguished between the two morphological environments. This result suggests that even *adult* speakers were storing these high-frequency nouns holistically. Evidence for lack of morphological decomposition in adult speech was surprising. Consequently, in an exploratory analysis, we hypothesized that any coarticulatory differences between adjacent phones may be washed out when averaging spectra over an entire phone. To test this, we next compared the difference (Euclidean distance) between the averaged spectral vectors taken from the *middle third* of each phone.

Figure 1: Predicting the Euclidean distance between Mel spectral vectors of [a] and [p].



We fit an additional model to predict the measurements from the middle third spectral vectors of each phone. The model again included random intercepts for **Word** and **Participant**. Now best model fit included the fixed effects of **Word duration** (β=3.90, t=2.24, p=.026, 97.5% CI=0.15, 7.35) and the interaction of **Morphological environment** with **Age** (β=1.96, t=2.26, p=0.024, 97.5% CI=0.28, 3.73) (Figure 1 under 'third phone model'; Figure 2).

Figure 2: Interquartile range of spectral distance between middle third of [a] and middle third of [p]. Black line represents median distance. Error bars represent 1.5x the IQR in each direction.

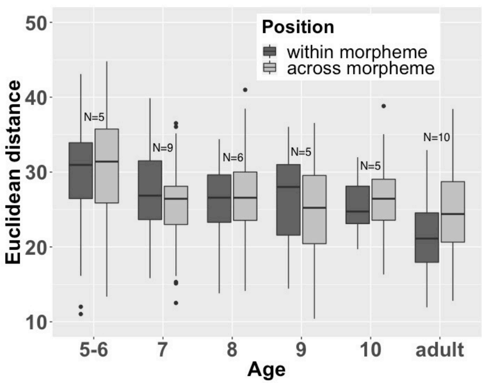


This interaction between age and morphological environment suggests that adults distinguish between [ap] sequences when they occur within morphemes versus between morphemes, but children do not. This could be because children do not decompose these high-frequency noun-suffix pairings, or at least not to the extent of adults.

Beyond differences between adults and children, we did not have a specific hypothesis about the age trajectory for this development. A factored variable of **Age group** – 5 child age bins and adults – did not improve model fit. However, Figure 3 displays a trend suggesting that 1) children do not distinguish coarticulatorily between the two morphological environments until age 10 and 2) the overall distance between phones decreases with age.

The trend for child participants to coarticulate more as they age is not consistent – note that the 10-year-olds pattern like adults but nine-year-olds appear to coarticulate *more* between morphemes than within. Furthermore, in a post-hoc analysis with 'adult' as the **Age group** reference level, only 7-year-olds reliably differed from adults (β=-3.54, *t*=-3.27, *p*=.001, 97.5% CI=-5.77, -1.45). There were no additional statistically significant differences between adults and any other child age group. This unclear pattern by age group may be due to the unbalanced/small sample sizes within each group. We hope to complement this analysis with more data collection on subsequent fieldtrips to disentangle the developmental trajectory.

Figure 3: Spectral distance between middle third of [a] and [p] across age groups. Black line represents median distance. Error bars represent 1.5x the IQR in each direction. Number of participants per age group listed above each box.



4. CONCLUSION

Speech does not unfold like pearls on a string. Coarticulation between segments is rampant and phonetic realization varies by context. Citing coarticulatory evidence, we demonstrated that in SBQ, a highly-agglutinating language, adult speakers have different coarticulatory patterns between versus within morphemes. This pattern is anticipated. Adult speakers have highly practiced motor routines, particularly within high-frequency words such as those we elicited. This experience results in well-rehearsed spectral transitions between the frequently co-occurring phones of a root morpheme. However, in adult speakers, this transitional routine is less practiced at the boundary of root morpheme and suffix. We take this as evidence that adult SBQ speakers have abstracted grammatical suffixes away from the original lexical contexts. Crucial to this argument is that child SBQ speakers do not distinguish between morphological environments. They coarticulate equally within and between morphemes. This is evidence that children store these high-frequency inflected nouns more holistically relative to adults. However, this conclusion came from an exploratory analysis taken from the middle third of phones. When we averaged over entire phones, neither adults nor children distinguished between environments. So while the exploratory analysis concludes that adults and not children distinguish between the morphological environments, this effect is washed away when averaging coarticulatory patterns over entire phones.

Our finding that children do not differentiate between morphological environments is novel evidence for an argument that has been made repeatedly in the child coarticulation literature: children represent language more holistically than adults [1], [5], [10]. However, we do not replicate reports that children coarticulate more than adults. Rather, coarticulation tends to *increase* with age (though we stress that the evidence by age group is a trend; there are too few participants per age group to definitively conclude). That coarticulation increases with age has been reported elsewhere [4], [5], [12], [20], though often the studies examined longer-distance coarticulation.

The differences between our child coarticulation conclusion and that of previous work could be the result of the spectral measurement that we employed, analysis of stops instead of fricatives, or measurement of coarticulation between syllables, not within. Regarding the last two concerns, we are limited by SBQ morpho-phonotactics – there are few fricative-initial suffixes that we could elicit with children.

However, there is another explanation for the result that children coarticulate more with age. What the child coarticulation literature has traditionally referred to as coarticulation likely does not refer to the planned, efficiency-driven process that characterizes adult speech [9], [25]. Children have less stable articulatory trajectories [21], so it would be surprising if they had greater coarticulatory control. Instead, previous research suggests that children coarticulate more than adults as a result of a lexical storage system where segments are underspecified. The result is that children do not discriminate between segments in production. This crucial distinction between planned, adult coarticulation and unskilled, child coarticulation is one that that few make (but see [20], [26]). However, our current findings – that children coarticulate similarly across morphological contexts while simultaneously coarticulating less than adults – supports a distinction between planned coarticulation and unskilled coarticulation. Planned coarticulation is a complex speech task that takes years of practice while unskilled coarticulation is the consequence of children's linguistic inexperience and holistic lexical storage.

5. REFERENCES

[1] Arnon, I. (2010). *Starting big: The role of multiword phrases* (Dissertation). Stanford University.

[2] Baayen, H. (1992). Quantitative aspects of morphological productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1991* (pp. 109–149). Kluwer Academic Publishers.

[3] Baker, R., Smith, R., & Hawkins, S. (2007). Phonetic differences between mis- and dis- in English prefixed and pseudo-prefixed words. In *Proceedings of the 16th ICPhS*. Saarbrücken.

[4] Barbier, G., Perrier, P., Ménard, L., Payan, Y., Tiede, M. K., & Perkell, J. S. (2013). Speech planning as an index of speech motor control maturity. In *Proceedings of Interspeech 2013*. Lyon, France.

[5] Barbier, G., Perrier, P., Ménard, L., Payan, Y., Tiede, M. K., & Perkell, J. S. (2015). Speech planning in 4-year-old children versus adults… In *Proceedings of Interspeech 2015*. Dresden, Germany.

[6] Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

[7] Berko, J. (1958). The Child's Learning of English Morphology. *Word*, *14*(2–3), 150–177.

[8] Boersma, P., & Weenik, D. (2018). Praat: doing phonetics by computer (Version 6.0.42).

[9] Bradlow, A. R. (2002). Confluent talker-and listener-oriented forces in clear speech production. In C. Gussenhoven & N. Warner (Eds.) *Laboratory Phonology*, *7*. (pp. 241-273). Berlin and New York: Mouton de Gruyter.

[10] Cho, T. (2001). Effects of Morpheme Boundaries on Intergestural Timing: Evidence from Korean. *Phonetica*, *58*(3), 129–162.

[11] Gerosa, M., Lee, S., Giuliani, D., & Narayanan, S. (2006). Analyzing Children's Speech: An Acoustic Study of Consonants and Consonant-Vowel Transition. In *2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings* (pp. 393–396). Toulouse, France: IEEE.

[12] Goffman, L., Smith, A., Heisler, L., & Ho, M. (2008). The Breadth of Coarticulatory Units in Children and Adults. *JSLHR*, *51*(6), 1424.

[13] Goodell, E. W., & Studdert-Kennedy, M. (1992). Acoustic Evidence for the Development of Gestural Coordination in the Speech of 2-Year-Olds: A Longitudinal Study. *Haskins Laboratories Status Report on Speech Research*, *SR-111/1123*, 63–88.

[14] Hay, J. (2003). *Causes and Consequences of Word Structure*. New York and London: Routledge.

[15] Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed-effects models. *Journal of Statistical Software*, *82*(13), 1–26.

[16] Lee-Kim, S.-I., Davidson, L., & Hwang, S. (2013). Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology*, *4*(2).

[17] Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *JSLHR*, *32*, 120–132.

[18] Pluymaekers, M., Ernestus, M., Baayen, R. H., & Booij, G. (2010). Morphological effects on fine phonetic detail: The case of Dutch -igheid. *Laboratory Phonology*, *10*, 511–531.

[19] R Core Team. (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

[20] Repp, B. H. (1986). Some observations on the development of anticipatory coarticulation. *JASA*, *79*(5), 1616–1619.

[21] Smith, A., & Goffman, L. (1998). Stability and Patterning of Speech Movement Sequences in Children and Adults. *JSLHR*, *41*(1), 18.

[22] Song, J. Y., Demuth, K., Evans, K., & Shattuck-Hufnagel, S. (2013). Durational cues to fricative codas in 2-year-olds' American English: Voicing and morphemic factors. *JASA, 133*(5), 2931–2946.

[23] Song, J. Y., Demuth, K., Shattuck-Hufnagel, S., & Ménard, L. (2013). The effects of coarticulation and morphological complexity on the production of English coda clusters: Acoustic and articulatory evidence from 2-year-olds and adults using ultrasound. *J. of Phonetics*, *41*(3–4), 281–295.

[24] Sugahara, M., & Turk, A. (2009). Durational correlates of English sublexical constituent structure. *Phonology*, *26*(03), 477.

[25] Whalen, D. H. (1990). Coarticulation is largely planned. *J. of Phonetics*, *18*, 3–35.

[26] Whiteside, S. P., & Hodgson, C. (2000). Speech patterns of children and adults elicited via a picture-naming task: An acoustic study. *Speech Communication*, *32*(4), 267–285.

[27] Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2011). Coarticulation as an Indicator of Speech Motor Control Development in Children: An Ultrasound Study. *Motor Control*, *15*(1), 118–140.

ACKNOWLEDGEMENTS

Research funded by Oswalt Documenting Endangered Language and U.C. Berkeley Student Technology Grants to the author. Thanks to Keith Johnson and Jan Edwards. Most importantly, thank you to the participating families.

---

[i] LOC=Locative

[ii] Eight participants knew their age but not birth date. Age averages do not include these participants.

[iii] Orthography is phonetically transparent.