An acoustic examination of the effects of word frequency and utterance position on 2-year-olds' production of 3rd person singular -*s*

Rachel M. Theodore[1], *Katherine Demuth*[2], *Stefanie Shattuck-Hufnagel*[3]

[1] Department of Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island, United States of America
[2] Department of Linguistics, Macquarie University, Sydney, Australia
[3] Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America

rachel_theodore@brown.edu, katherine.demuth@mq.edu.au, sshuf@mit.edu

Abstract

Some variability in children's productions of grammatical morphemes reflects contextual factors. Transcription has shown that 3rd person singular –*s* production is more robust utterance-finally compared to utterance-medially, and in simple versus complex codas. Acoustic analyses were used to examine further the influence of utterance position, and to examine whether word frequency affects morpheme production. The speech of 14 children (mean age = 2;3) was obtained using an elicited-production task. Results replicated the positional effect, but showed that morphemes were equally produced in low versus high frequency verbs, which suggests that contextual and lexical factors contribute differentially to morphological development.

Index Terms: language acquisition, acoustic analysis, morpheme production

1. Introduction

Children's early productions are highly variable, even for a given speaker. Such variability is present for productions of monomorphemic speech segments [1], as well as for early productions of grammatical morphemes [2]. This variability has been taken as evidence that early grammars are syntactically incomplete [3]. However, recent findings have shown that much of the early variability in morpheme production is systematically governed by contextual and semantic factors. For example, the past tense morpheme is produced earlier for accomplishment verbs compared to activity verbs, and the plural morpheme is more robustly produced when labeling a set of objects that are similar versus distinct [4]. Grammatical morphemes are also produced more reliably when they occur as part of a trochaic foot compared to when they are unfooted [5] or as part of a simple coda consonant rather than a complex coda cluster [6]. Such findings challenge the view that variability in production of grammatical morphemes reflects impoverished syntactic representations, and raise the possibility that variable within-speaker production of grammatical morphemes may reflect interactions between syntactic and other levels of language representation and planning.

This view has been formalized in the prosodic licensing hypothesis of Demuth and McCullough [7], which predicts that children are more likely to produce morphemes in phonologically simple, or unmarked, contexts. Recent support for this hypothesis comes from Song et al. [8], who examined the production of 3rd person singular –*s* in 2-year-old children. Their results, drawn from perceptual analysis of both corpus and experimental data, were two-fold. First, production of 3rd person singular –*s* was more robust for verbs with simple codas (e.g., *cries*) compared to verbs with cluster codas (e.g., *drives*). Second, production of 3rd person singular –*s* was more robust when the verb was produced utterance-finally compared to utterance-medially (e.g., *There he cries* vs. *He cries now*). Thus, their results showed that morpheme production was better in phonologically less difficult contexts, both at the level of the syllable/word (i.e., coda complexity) and the phrase (i.e., utterance position).

In addition to phonological influences, lexical factors have been shown to affect the production of grammatical morphemes. For examples, plural –*s* in typically-developing children has been shown to be more often produced in high frequency nouns compared to low frequency nouns [9] and similar findings have been observed for children with specific language impairment [10]. However, word frequency was not found to affect production 3rd person singular –*s* in the corpus analysis performed by Song et al. [8]. The discrepancy between these findings may reflect myriad differences between the studies, including the age of children being examined, the degree to which they were typically developing versus language-impaired, as well as the nature of corpus studies versus experimental manipulations. Given the importance of specifying the degree to which both phonological and lexical factors interact with children's early grammatical productions, here we explore further these influences on production of 3rd person singular –*s*. The current work therefore extends previous findings in two ways. First, instead of perceptual transcription, we examine the acoustics of children's speech. Recent studies have shown that transcription can be influenced by many factors including the perceived age of the speaker, the phonetic expertise of the transcriber, whether or not the speaker and transcriber share the same native language, as well as whether or not the transcriber is given one or multiple categories in which to designate a transcription [11]. Second, we examine the effects of utterance position and word frequency in more carefully controlled phonetic environments than in previous studies.

2. Methods

2.1. Subjects

The subjects were 14 children from the Providence, Rhode Island community. All were full-term 2-year-olds (6 girls, 8 boys) from monolingual English-speaking homes, and ranged in age from 2;0 - 2;6 (mean 2;3). All were healthy on the day of testing and had typically-developing speech and language skills according to parental report. MacArthur CDI percentile scores on vocabulary size ranged from 25-90+, with a mean of

76. An additional seven children were not included in the analyses. Five were excluded because they did not speak during the experiment, and two were excluded because they did not meet the criterion for inclusion, which was to repeat at least 12 of the 16 prompts presented during the experiment. This attrition rate is consistent with previous experiments using similar tasks with children in this age group [8].

2.2. Stimuli

Eight target verbs were selected for use in the experiment, half of which were selected to be "high frequency" items and half of which were selected to be "low frequency" items. Frequency was calculated on verbs inflected with the 3rd person singular morpheme, and was determined using three metrics, all of which are available on the CHILDES database [13]. The first metric consisted of frequency counts of many corpora of child-directed speech [12]. The second and third metrics consists of frequency counts derived from the Providence Corpus, one for the adults and the other for the children. The sum frequency across these three metrics was used to categorize the target items into low and high frequency categories. The mean frequency of the two categories was highly distinct ($mean_{low} = 26.75$, $mean_{high} = 222.75$), and the frequency distributions of the two categories did not overlap ($range_{low} = 21 – 35$, $range_{high} = 72 – 592$).

An attempt was made to control the lexical and segmental characteristics of the target verbs. All consisted of CV(C)–*s* syllable structure, with the final consonant consisting of a voiceless stop. Each target verb was embedded in two sentences, one in which it appeared utterance-medially and one in which it appeared utterance-finally (see Table 1). To control for length, all stimulus sentences consisted of three monosyllabic words. To control for articulatory and phonological influences at the phrase level, the utterance-medial sentences consisted of the target verb followed by a word that began with a labial stop consonant.

Table 1. *Stimulus sentences for the high and low frequency targets in medial and final utterance position.*

		Utterance position	
Frequency	Target	Medial	Final
Low	*hits*	She <u>hits</u> balls.	Now she <u>sits</u>.
	fights	He <u>fights</u> bears.	Now he <u>fights</u>.
	cooks	He <u>cooks</u> peas.	There he <u>cooks</u>.
	bakes	She <u>bakes</u> pies.	There she <u>bakes</u>.
High	*sits*	He <u>sits</u> back.	There he <u>sits</u>.
	bites	She <u>bites</u> pears.	There she <u>bites</u>.
	talks	She <u>talks</u> back.	Now she <u>talks</u>.
	takes	He <u>takes</u> books.	Now he <u>takes</u>.

Eight "movies" were created (one for each target verb) to serve as visual prompts. All movies were cartoon style and were pre-screened to be of similar size and interest, as well as to be an accurate representation of the target verb.

In addition to the visual prompts, auditory prompts were created. A female native speaker of American English was recorded producing the sixteen sentences in infant-directed speech register. The recording session took place in a sound-attenuated booth. Speech was recorded to computer at a sampling frequency of 44.1 KHz and 16-bit quantization via a microphone connected to a pre-amplifier. Each sentence was excised into separate files using the Praat software [14].

For each sentence, three acoustic measurements were performed using Praat. First, we measured the duration of the entire sentence, calculated as the latency between the onset and offset of vocal energy present in the utterance. Second, we measured the duration of the target noun as the difference between the onset of closure for the initial stop consonant and the offset of high frequency frication energy associated with the 3rd person singular *–s*. Third, we measured the duration of the 3rd person singular *–s*, taken as the difference between the onset and offset of high frequency aperiodic noise associated with fricative production. These measurements were performed in order to ensure that the overall duration of the stimulus sentences was comparable for the high and low frequency items in both utterance positions. Moreover, we wanted to confirm that the duration of the target verb and duration of the 3rd person singular morpheme was equivalent across the low and high frequency verbs, but longer in utterance-final compared to utterance-medial position. ANOVA was used to confirm the patterns described above, and these results indicate that the selected stimuli are comparable to those used for the 3rd person singular experiments reported in Song et al. [8].

2.3. Procedure

The child was invited into a sound-attenuated room with a parent to "play a game" with the experimenter. The room was equipped with two lavalier microphones (Audio-technica 700 Series) connected to a computer in an adjoining room via an MBox 2 Audio Interface (Digidesign). In most cases, the microphones were placed on the table near the child in order to best capture his or her speech; in a few cases, one of the microphones was attached to the collar of the child's shirt. Following a brief warm-up period where the child was asked to repeat what the computer said, the child was directed to face the computer in order for the game to begin.

On each trial, a picture of the target verb appeared on the monitor along with the auditory prompt and a movie depicting the action of the target verb, and the child was directed to repeat the prompt. Five attempts were allowed before moving to the next trial. The child was encouraged with praise and stickers. Following the completion of the sixteen trials, parents were asked to fill out the short form of the MacArthur CDI in order to estimate the child's vocabulary size. The entire procedure took approximately thirty minutes.

2.4. Acoustic analysis

Each utterance was excised using Praat and saved to an individual file for subsequent acoustic coding. Following conventions established by Shattuck-Hufnagel et al. [15], each utterance was coded for an assortment of acoustic cues to the distinctive features of the coda segments, three of which are reported in the current paper. This method is based on Steven's feature-cue-based model, which proposes that a given feature contrast may be signaled by a number of different acoustic cues, and that the precise set of cues that a speaker employs may vary depending on the other features in the feature bundle (i.e. the phonemic segment), as well as on the segmental and structural context in which the feature occurs [16]. This model separates the acoustic cues in the speech signal into acoustic landmarks, i.e., robustly-detectable abrupt changes in the acoustic signal, and an additional set of cues which are found in the vicinity of the landmarks. The landmarks provide information about the articulator-free features that correspond (roughly) to manner of articulation, and the additional cues provide information about articulator-bound features such as voicing and place of articulation.

All cues were identified by combining visual inspection of the waveform and spectrogram with listening. Figure 1 shows a representative waveform and spectrogram illustrating the two acoustic cues of interest. The first cue, which served as the acoustic metric of 3rd person singular –*s* production, was the presence of high frequency aperiodic noise following either periodicity for the vowel or the release burst and associated noise following stop production. Amplitude and duration of this region of noise were used to differentiate noise produced following the stop release burst. If the acoustic analyses pattern the same as earlier transcription analyses, then we should observe the increased presence of high frequency, aperiodic noise in targets produced utterance-finally compared to utterance-medially. Furthermore, if word frequency influences morpheme production as does the phonological factor of coda complexity, then we should also observe the increased presence of high frequency aperiodic noise in high frequency compared to low frequency verbs.

The second cue, release burst, is associated with stop consonant production. A release burst results from the release of pressure generated by occluding the vocal tract for stop production while air continues to flow through the vocal folds into the mouth, and acoustically manifests as a sudden spike of transient energy. Measuring this cue allowed us to examine in detail the degree to which cluster codas (stop + –*s*) were potentially reduced to singleton codas (–*s* only or stop only).

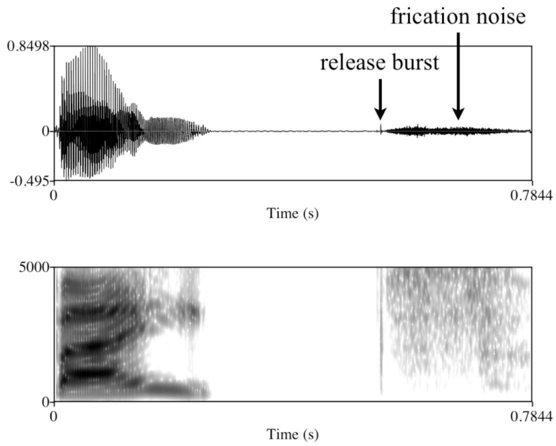

Figure 1: *Representative waveform and spectrogram showing acoustic landmarks for the target "bites", which was excised from the utterance-final sentence used as an auditory prompt.*

3. Results

3.1. Effects of utterance position and word frequency on 3rd person singular –*s*

The presence or absence of high frequency aperiodic noise was used to categorize each utterance as containing 3rd person singular –*s* or not containing 3rd person singular –*s,* respectively. For each child, the proportion of tokens that contained the morpheme (as indicated by the presence of frication noise) was calculated for low and high frequency verbs in utterance-medial and utterance-final positions by collapsing across the four tokens of each type. Figure 2 shows mean morpheme production across the 14 children (error bars indicate standard error of the mean). In order to examine the statistical significance of children's performance, mean 3rd person singular –*s* production across children was submitted to repeated-measures ANOVA with the factors of verb frequency (high vs. low) and utterance position (medial vs. final).

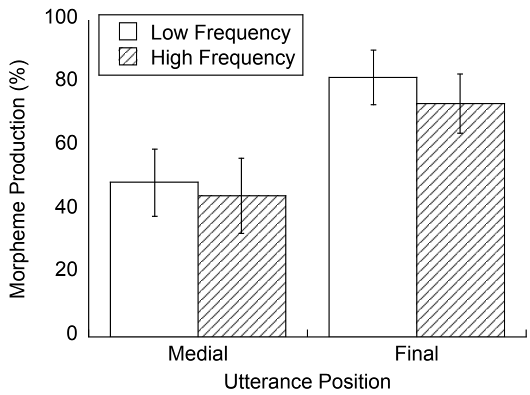

Figure 2: *Percentage of high and low frequency verb tokens produced with the 3rd person singular morpheme in utterance medial and utterance final positions.*

As expected, results of ANOVA revealed a main effect of position [$F(1,13) = 11.88$, $p = .004$], in that the morpheme was produced in a greater proportion of utterance-final tokens (78.0%) compared to utterance-medial tokens (46.7%). However, there was no main effect of verb frequency, indicating that morpheme production was equivalent across both low and high frequency verbs [$F(1,13) = 1.07$, $p = .319$]. Moreover, there was no interaction between utterance position and coda complexity [$F(1,13) = .006$, $p = .741$]. Thus, these findings are in accord with Song et al. [8] who found that production of 3rd person singular –s is influenced by the position of the verb in the utterance, with better production in utterance-final compared to utterance-medial position. Contra other reports, however, we found no evidence that word frequency influences production of 3rd person singular, despite using more fine-grained acoustic analysis compared to perceptual transcription. These findings led us to conduct an additional analysis of the coda clusters in order to determine if word frequency would affect the degree to which the clusters were simplified to singleton consonants, even though it did not affect the presence of the morpheme in the cluster.

3.2. Effects of utterance position and word frequency on coda cluster production

A second analysis was conducted in order to examine more closely the acoustic realization of the elicited codas. Recall that for 3rd person singular –*s*, morpheme production is more robust in simple compared to complex codas [8]. This finding may reflect the fact that cluster codas present an articulatory challenge to the developing system compared to singleton codas. In this analysis we explored the hypothesis that effects of word frequency on morpheme production may be realized when considering production of the entire coda. Thus, each utterance was categorized as containing a cluster coda if the utterance contained a release burst and high frequency, aperiodic frication noise (with frication noise being distinguished from post-release noise by both duration and amplitude), and was categorized as not containing a cluster if both of these cues were not present.

Figure 3 shows the mean cluster production across the 14 children (error bars indicate standard error of the mean). A repeated-measured ANOVA with the factors of verb frequency (high vs. low) and utterance position (medial vs. final) was used to examine the statistical significance of the children's performance. The results of the ANOVA paralleled those observed for production the morpheme. Specifically, more clusters were produced in utterance-final compared to utterance-medial position, with equivalent cluster production across low and high frequency verbs. Furthermore, there was no interaction between utterance position and verb frequency. Taken together, the results from both analyses indicate that for the 2-year-old children in the present sample, verb frequency mediates neither the presence of 3rd person singular *–s*, nor the degree to which cluster codas are produced.

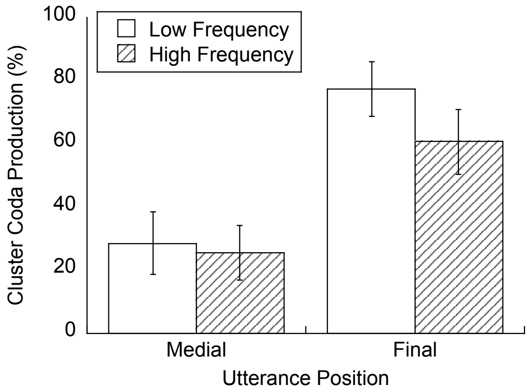

Figure 3: *Percentage of high and low frequency verb tokens produced with a cluster coda in utterance medial and utterance final positions.*

4. Discussion

As reviewed previously, children's early productions of grammatical morphemes are highly variable. Given the fact that some of this variability systematically reflects both semantic and phonological factors, variability in productions is not solely the consequence of incomplete syntactic representations. Rather, such variability likely reflects interactions between syntactic and other representational levels in the developing language system. Specifying precisely how these different levels interact during language production is indeed necessary in order to construct models of the acquisition process. To this end, here we examined the influences of word frequency and utterance position on production of 3rd person singular –s. Morpheme production was enhanced utterance-finally compared to utterance-medially, but was equal across high and low frequency verbs. Furthermore, verb frequency did not influence production of coda production in general, as reflected in the equal production of the coda clusters across frequency types.

Recall that earlier work has shown that word frequency can mediate production of plural –s [9, 10]. One possible explanation for the lack of a frequency effect presently may lie in the nature of the syntactic and semantic differences between plural and verb morphology. In line with this hypothesis, recent findings have shown that though coda complexity influences production of 3rd person singular *–s*, it does not influence production of plural *–s* [17]. Such findings suggest that the nature of the morphological representation proper influences the degree to which other levels of representation will impede or facilitate production of grammatical morphemes. In addition, in light of studies showing influences of phonotactic probability (as a metric of lexical density) on children's productions [18], it should be noted that the absence of lexical frequency effects may not reflect a lack of lexical influences on 3rd person singular production; lexical properties may indeed influence morpheme production, but perhaps via segmental relationships within lexical entries rather than via summed frequency of exposure to a particular lexical entry *per se*. Future work is aimed as examining this issue.

5. Conclusions

Production of 3rd person singular *–s* in 2-year old children was more robust utterance-finally compared to utterance-medially, but equal in low versus high frequency verbs. This suggests that phonological and lexical factors contribute differentially to morphological development.

6. Acknowledgements

This research was supported by NIH Grant No. R01 HD 057606.

7. References

[1] Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis, and coda licensing in the early acquisition of English. *Language & Speech, 49,* 137-174.

[2] Bloom, L. (1970). Language development: Form and function in emerging grammars. Cambridge, MA: MIT Press.

[3] Marcus, G. F., Ullman, M., Pinker, S., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development, 57(4)*, Serial No. 228.

[4] Zapf, J. A. & Smith, L. B. (2008). Meaning matters in children's plural productions. *Cognition, 108,* 466-476.

[5] Gerken, L. A. (1996). Prosodic structure in young children's language production. *Language, 72,* 683-712.

[6] Marshall, C. & van der Lely, H. (2007). The impact of phonological complexity on past tense inflection in children with Grammatical-SLI. *Advances in Speech Language Pathology, 9,* 191-203.

[7] Demuth, K., & McCullough, E. (2009). The prosodic (re)organization of children's early English articles. *Journal of Child Language, 36,* 173-200.

[8] Song, J., Sundara, M., & Demuth, K. (2009). Phonological constraints on children's production of English third person singular *–s*. *Journal of Speech, Language, and Hearing Research, 52,* 623-642.

[9] Zapf, J. (2004). Frequency in the input and children's mastery of the regular English plural. In A. Brugos, L. Micciulla, & C. E. Smith (Eds.), *Proceedings of the 28th Annual Boston University Conference on Language Development* (Vol. 2, pp. 669–680). Somerville, MA: Cascadilla.

[10] Oetting, J. B., & Rice, M. (1993). Plural acquisition in children with specific language impairment. *Journal of Speech and Hearing Research, 36,* 1241–1253.

[11] Munson, B., Edwards, J., Schellinger, S., Beckman, M. E., & Meyer, M. (2010). Deconstructing phonetic transcription: Language-specificity, covert contrast, perceptual bias, and an extraterrestrial view of vox humana. *Clinical Linguistics and Phonetics, 24,* 245-260.

[12] Li, P., & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect.* Berlin & New York: Mouton de Gruyter.

[13] MacWhinney, B. (2000). The CHILDES project. Mahwah, NJ: Erlbaum.

[14] Boersma, P. & Weenink, D. (2010). Praat: doing phonetics by computer (Version 5.1.25) [Computer program]. Retrieved January 20, 2010, from http://www.praat.org/.

[15] Shattuck-Hufnagel, S., Demuth, K., Hanson, H. & Stevens, K. N. (In Press). Acoustic cues to stop-coda voicing contrasts in the speech of American English 2-3-year-olds. In G. N. Clements & R. Ridouane (Eds.). *Where do feature comes from? The nature and sources of phonological primitives.* Elsevier: North-Holland Linguistic Series.

[16] Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America, 4,* 1872-1891.

[17] Theodore, R. M., Demuth, K., & Shattuck-Hufnagel, S. (Under review). Acoustic evidence for position and complexity effects on children's production of plural *–s*.

[18] Zamuner, T. S. (2009). Phonological probabilities at the onset of language development: Speech production and word position. *Journal of Speech, Language, and Hearing Research, 52*, 49-60.