



The acquisition of phonology based on input: a closer look at the relation of cross-linguistic and child language data[☆]

Tania S. Zamuner^{a,*}, LouAnn Gerken^b,
Michael Hammond^c

^a*Interfaculty Research Unit for Language and Speech (IWTS), University of Nijmegen,
Wundtlaan 1, 6525 XD Nijmegen, The Netherlands*

^b*Department of Psychology, University of Arizona, Tucson, AZ 85721, USA*

^c*Department of Linguistics, University of Arizona, Tucson, AZ 85721, USA*

Received 30 October 2003; received in revised form 16 June 2004; accepted 16 June 2004

Available online 1 August 2004

Abstract

Parallels between cross-linguistic and child language data have been used to support a theory of language development in which acquisition is mediated by universal grammar (Universal Grammar Hypothesis—UGH). However, structures that are frequent across languages are also typically the most frequent within a specific language. This confounding of cross-linguistic and language-specific data is consistent with the hypothesis that children acquire the grammar of the specific languages to which they are exposed, based on a less constrained acquisition mechanism (Specific Language Grammar Hypothesis—SLGH). These two theories of acquisition are contrasted in an examination of English-speaking children's acquisition of codas. Predictions for the UGH were based on cross-linguistic patterns and on frequency analyses of codas from 35 languages. Results showed that languages prefer coronal and sonorant codas; however, children's productions did not favor these codas. Predictions for the SLGH were established on the frequency of English codas, and significant correlations were found between children's coda productions and the frequency of English codas.

[☆] This research was supported by NSERC PGSB awarded to T.S. Zamuner, by NSF grant #SBR969607 awarded to LouAnn Gerken, and by NWO grant 360-70-100 awarded to Paula Fikkert and Rene Kager. We are thankful for the assistance and comments from Paula Fikkert, Diane Ohala, Maryam Sadrzadeh, Joseph Stemberger, Erik Jan van der Torre, and an anonymous reviewer.

* Corresponding author. Tel.: +31 24 361 29 89; fax: +31 24 361 19 72.

E-mail address: t.zamuner@mpi.nl (T.S. Zamuner).

Using this approach, children's coda production is best characterized with respect to frequently occurring properties of the input, which serve to organize children's linguistic representations.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Acquisition; Phonology; Universal grammar; Input

1. Introduction

Theories of human language and its acquisition by children often take as crucial data the noted parallel between cross-linguistic markedness and child language. The primary account of this relationship relies on the Chomskian notion of innate universal grammar (UG).¹ We refer to this theory of grammar acquisition as the Universal Grammar Hypothesis (UGH). However, as we will demonstrate in more detail below, the data taken to support this account are often partially confounded with the input to language acquisition. In particular, the distributions of sound patterns in a single language often encode what is argued to be innately provided by UG. The partial confounding of cross-linguistic and language-specific data allow for an alternative to the UGH, which we will refer to as the Specific Language Grammar Hypothesis (SLGH). On this view, children acquire the grammar of the specific languages to which they are exposed by applying a less constrained acquisition mechanism to patterns in the input. We will attempt to contrast these two hypotheses using the acquisition of codas by children acquiring English.

Let us first consider the role that cross-linguistic data have played in the theory of UG. The discussion will center on two different characterizations of UG that can be found in the literature. The first is that UG contains the set of properties present in all languages. This set comprises unrestricted or absolute universals. Unrestricted universals refer exclusively to the necessary properties that define all languages or what are also called the “unmarked” properties of languages. For example, all languages have CV syllables (Spencer, 1996: 82); moreover, there are no languages without CV syllables. Unrestricted universals are by definition based on cross-linguistic research (Hammond et al., 1988). The second characterization of UG is that it defines the limits of human language and the extent to which languages can vary. Thus, UG contains and is defined by what is unmarked and marked in language. The term “unmarked” refers to the properties of language that are common and frequent (which can be different from what is frequent in any one language), whereas “marked” defines properties that are rarer and less frequent. Various methods are used to determine what is unmarked and marked in languages, such as being based on restricted universals (also called implicational universals or typological universals, Greenberg, 1963, 1966), language processes (Greenberg, 1966), language change, and based on patterns seen in child language and aphasia (Blumstein, 1973; Jakobson, 1941/1968). One

¹ The view that language acquisition is mediated by UG stems from Jakobson's theory of acquisition within the framework of the Prague School of Linguistics (Jakobson, 1941/1968). For Jakobson, the same laws that govern the phonemic systems of adult languages also govern children's acquisition of contrasts. Thus, he made the important observation that there are similarities between adult language and child language. He further noted that the order of acquisition can be predicted from cross-linguistic implicational universals.

that is key in the current discussion concerns what are alternately referred to as statistical, substantive or distributional universals (Greenberg, 1954, 1966, 1974; Hammond et al., 1988; Trubetzkoy, 1939/1969; Zipf, 1935/1965). These types of generalizations are not absolute, but rather, they describe languages' tendencies to pattern in a particular way. Not only is markedness connected to frequency across languages, but it is also encoded in the frequency of sound patterns within a single language. That is, unmarked segments tend to be frequent within the distribution of a single language's sound patterns, and marked segments are typically less frequent (e.g., Ferguson, 1963; Greenberg, 1966; Trubetzkoy, 1931, 1939/1969; Zipf, 1935/1965). For example, CV syllables are preferred cross-linguistically, and although English allows a large number of syllable types, CV syllables are the most frequent in English words.

Although the characterization of UG that is relevant for language acquisition is subject to debate, what is constant across all theories is that UG minimally provides to children what is unmarked in language, as part of children's innate genetic endowment. The Universal Grammar Hypothesis predicts that child language data should conform to what is unmarked across languages. A frequently cited piece of evidence supporting this prediction concerns syllable structure: The universally unmarked syllable shape is CV, such that all languages permit CV syllables (Blevins, 1995; Clements and Keyser, 1983; Greenberg, 1978). The least marked languages permit only this syllable type and do not permit codas, e.g., Arabela, Kikuyu and Twi. Interestingly, children's early word productions are largely characterized by this same open syllable shape. For example, a child attempting to produce the final /n/ in "man", might delete /n/ and produce [mæ], [me], or [mʌ] (Demuth and Fee, 1995). From this example, we can see that the child's productions conform to an onset-peak or CV shape. One interpretation of this phenomenon is that final consonant deletion results from innate pressures to conform words to the preferred CV shape. Thus, these errors have been traditionally interpreted as a reflection of innate UG (e.g., Demuth and Fee, 1995; Faingold, 1990; Levelt et al., 2000; Mowrer and Burger, 1991; Ohala, 1996; Vihman and Ferguson, 1987). Within generative theories of linguistics, the unmarked properties of language or the core grammar (Chomsky, 1981) comprise children's initial hypotheses about language. Thus, children's initial CV productions are explained by appealing to the CV syllable as the unmarked syllable type or state. As children's grammars develop, they move away from the core grammar towards a language-specific grammar. Consequently, children produce more marked syllable structures.²

There are two important points to note from the preceding discussion. First, it is commonly argued that UG is evidenced by cross-linguistic markedness effects. Second, what is frequent across languages is often frequent in a particular language. The last statement suggests an alternative to the UG hypothesis for language acquisition, which is

² Children do not necessarily acquire unmarked structures first and marked structures last (Hyams, 1986). For instance, although /l/ and /r/ are preferred consonants cross-linguistically, these sounds are difficult for children to perceive and produce (Poole, 1934; Welleman et al., 1931). Given that children exhibit perceptual and production constraints unlike adults, this raises an interesting question of how to reconcile these differences. Factors other than UG are evidenced in acquisition, and domains can interact with one another. These are complex issues that need to be addressed. As a starting point, however, we take the standard approach assumed in the acquisition literature, where adult language patterns are seen as reflecting the properties of UG.

that language acquisition is best described with respect to the patterns in the input or ambient language. We refer to the latter view as the Specific Language Grammar Hypothesis. This hypothesis is motivated by the large body of research that illustrates that child language reflects language-specific input at an early age. For example, high frequency sounds in the ambient language are more likely to be produced by children than are low frequency sounds. Research looking at the acquisition of codas demonstrates that children learning English produce codas before children learning Spanish (Demuth, 2001; Gennari and Demuth, 1997). (See also Fikkert and Freitas, 1997 for Dutch and Portuguese.) It is suggested that this production difference can be attributed to differences in coda frequencies in the two languages; codas are more frequent in English than in Spanish (Roark and Demuth, 2000). Other research looking at the acquisition of identical phonemes in different languages has shown that acquisition rates are related to the frequency in the respective languages (e.g., Ingram, 1988; Macken, 1995; Pye et al., 1987).

To summarize, cross-linguistic markedness data, which have traditionally been viewed as evidence for the UGH, are also potentially consistent with predictions based on the SLGH. In the domain of phonology, the UGH predicts that children should initially produce those sound patterns that are unmarked or frequent across languages before those patterns that are marked or infrequent. The SLGH predicts that children should initially produce the more frequently occurring sound patterns in their ambient language before producing the less frequent ones. As noted earlier, what is frequent across languages is typically also frequent within a specific language, making it appear difficult to contrast the two hypotheses. However, the patterns of cross-linguistic and language-specific data do not necessarily match perfectly. Therefore, if we look at the areas of mismatch and compare these to child language data, we can ask whether the patterns seen in acquisition are more consistent with cross-linguistic or language-specific data.

The domain in which we will compare the UGH and SLGH is word-final codas. This domain is well suited to our inquiry, because coda distributions are restricted both across languages and within languages, and children's initial productions of codas are also restricted. For purposes of the current investigation, we will define a coda as a syllable- or word-final consonant, such as the "k" in "puck."³ The data considered throughout the study are restricted to word-final codas in monosyllabic CVC words for two primary reasons. First, there is no consensus as to how intervocalic consonants are syllabified; different linguistic theories predict different syllabifications for medial consonants (e.g.,

³ In government phonology, word-final consonants are syllabified as onsets of empty-headed syllables (Harris, 1997; Kaye et al., 1990; Piggott, 1991, 1999). Under this assumption, Goad (1997) and Goad and Brannen (in press) argue that children's first productions of final consonants are initially syllabified as onsets to account for the phonetic properties of these consonants (aspiration, length and homorganic nasal release). Word-final consonants have also been syllabified as part of an appendix and as extraprosodic (Booij and Rubach, 1984; Borowsky, 1986; Halle and Vergnaud, 1980; Itô, 1986; McCarthy, 1979). Although these alternative representations are not necessarily incompatible with the research presented here, these devices do not capture the probabilistic patterns of word-final consonants. The goal of this study is not to qualify or define a specific prosodic representation. Rather, we illustrate that there are consistent cross-language preferences in word-final position. This research also does not explore the acquisition of rhyme structures, as explored in Demuth and Fee (1995), Fikkert (1994) and Kehoe and Stoel-Gammon (2001). See Kehoe and Stoel-Gammon for a recent review of this literature.

Hammond, 1997; Kahn, 1976; Selkirk, 1982). In addition, native speaker's intuitions of how these consonants are syllabified are affected by vowel quality, consonant type, stress, spelling, and morphology (e.g., Derwing, 1992; Fallows, 1981; Treiman and Danis, 1988; Zamuner and Ohala, 1999). Thus, determining the syllabic constituency of medial consonants is a complicated matter, and different choices about constituency are likely to influence any measurement of cross-linguistic or language-specific coda frequency. Second, the set of possible codas, the frequency of codas, and children's production of codas are all affected by prosodic environment (Franke, 1912; Hammond, 1999; Kirk and Demuth, 2004; Zamuner, 2003; Zamuner and Gerken, 1998). To circumvent these problems, the analyses are restricted codas to final position in monosyllabic CVC words.⁴ We explore the place and sonority of word-final codas in the domains of cross-linguistic markedness, the distributions of codas in English and in coda acquisition in English. We begin with an examination of codas across languages.

2. Codas in cross-linguistic data

The traditional basis of determining markedness comes from restricted universals, language processes, and language change. Based on this formulation, there is wide spread agreement that codas across languages tend to be coronal (Paradis and Prunet, 1991) and sonorant (Clements, 1990: 303), and much of the data for these claims are cross-linguistic in nature. However, the existing cross-linguistic data do not allow a direct comparison to adult English because English allows codas with all places of articulation, and English allows both sonorant and obstruent codas. Importantly, none of the reported cross-linguistic data concern only codas of CVC words, which means these cross-linguistic data are not directly comparable to adult English or child language data. Therefore, in order to create a firm foundation for comparing among the three types of data of interest, we conducted our own analysis of previously published research and data collection of lexicons of CVC words from 35 languages (see Appendix A for a list of the languages and Zamuner, 2003 for a full description). To obtain the CVC data, grammars and dictionaries were scanned for all or a random subset of CVC words. The set of possible codas and number of word-final codas for each language were then tabulated with respect to place and sonority. The size of the lexicons ranged from 33 to 1,153 words ($M = 242.8$, median = 148). The size of the lexicons reflects both the completeness of the grammars used, and the number of codas in the language. For example, although the Latvian lexicon had only 64 CVC words, these words were collected from two grammars and a dictionary with approximately 37,638 words. The data presented here are based on type word counts of lexicons. It is assumed that the proportions of codas that exist in the samples are representative of the proportions of codas that exist in the individual languages as a whole.

⁴ By restricting analyses to word-final position in monosyllabic words, many potential confounds are avoided, such as those arising from positional markedness. Because of this restriction, there are a number of other interesting factors that will not be addressed, which are worth investigating. For example, word-final consonants at the ends of unstressed syllables are restricted to a smaller set and less marked consonants. Inventories of codas in word-final position relative to word-medial position are also not considered, which is shown by Piggott (1999) to be an important factor to consider when looking at the distribution of codas across languages.

The 35 languages were chosen to provide data on CVC words from a variety of languages and sub-families. It is assumed that the sample is random.

The motivation for looking at statistical preferences for codas across languages stems from the observation that cross-linguistic markedness is reflected in the text frequency of languages. Unmarked segments are generally more frequent than marked segments (e.g., Ferguson, 1963; Greenberg, 1966; Trubetzkoy, 1931, 1939/1969; Zipf, 1935/1965). The assumption is that a cross-linguistic statistical analysis of word-final codas will allow for a further clarification of what is unmarked and marked in this position. If tendencies are found across languages, this is taken to reflect cross-linguistic markedness (combined with previous and more traditional ways of establishing cross-linguistic markedness). Because cross-linguistic markedness patterns are interpreted here as evidence for UG, an analysis of these patterns will subsequently allow for a prediction of what patterns are expected to emerge in children's productions of codas. The Prague School of Linguistics emphasized the pertinence of looking at the distribution of a language's sounds both in terms of the "ratio of these figures to the figures of occurrences theoretically expected on the basis of combinatory rules" (Trubetzkoy, 1939/1969: 257) and in terms of absolute numbers of occurrences. Therefore, two different frequency analyses were done on the data. The first analysis will be referred to as an Expected Frequency Analysis (EFA) and the second will be referred to as an Actual Frequency Analysis (AFA).

Expected Frequency Analyses were done to determine whether the words of a language contain a specific phonological element more often than expected by chance. In this case, chance is based on the number of codas with a specific phonological feature (N_{PFC}), divided by the number of possible codas types in the language (N_{C}), multiplied by the number of words in the lexicon (N_{L}). The result gives the number of words one would expect to find with a specific phonological feature in the lexicon. The result of the EFA is then compared to the number of words in the language that do, in fact, have the coda type in question. To illustrate, take Dutch (Baayen et al., 1993), which has eleven codas in the inventory (p, f, m, t, s, n, r, l, k, x, ŋ); five of these bear the feature coronal. By chance, one would expect a proportion of .45 coronal codas (5 coronal codas divided by 11 codas in the inventory) in the total CVC lexicon (1,153 words). Thus, it is expected that 524 words should end in coronal codas (.45 by 1,153). In actuality, 669 words end in coronal codas. By considering expected frequencies, we can control for the fact that languages have more coronal codas in their inventories than codas with other places of articulation. An EFA was computed for each language and data from the 35 languages were combined to determine the overall cross-linguistic effect, using a chi-square goodness-of-fit test.

Actual Frequency Analyses were done to determine whether there were a greater number of codas with a certain phonological feature across the 35 languages. In this sense, the AFA simply measures which are the most frequent codas in a language. To illustrate this type of analysis, consider again place of articulation in Dutch: 669 of the lexicon's words end in coronals, whereas 484 of the words end in labials and dorsals combined. It would also be possible to separate labial and dorsal codas from coronal codas, however, combining them provides a more generous and less biased estimate. The reason for this is that there are many coronal codas in languages' inventories, but comparatively few labial

and dorsal codas. An AFA was performed on each language. To determine whether there were significantly more codas with a specific phonological feature than other features, a t-test was performed by collapsing the data across the 35 languages to provide the general cross-linguistic pattern.

It is generally claimed that coronal consonants are unmarked (see Maddieson, 1984; Paradis and Prunet, 1991).⁵ This claim is based on traditional characterizations of universal grammar and markedness, e.g., based on restricted universals, language processes, and language change. Consistent with this claim, cross-linguistic frequency analyses of word-final codas were examined with respect to their place of articulation as either labial, coronal, or dorsal. An EFA across the 35 languages revealed that there were more coronal codas than expected, $\chi^2(34, N = 7906) = 209.29, P < .001$. To determine whether the pattern of results was significant for the individual languages, a chi-square goodness-of-fit test was run on the data from each language. Results indicated that 21 of the 35 languages had significantly more coronal codas than expected, and four languages had fewer coronal codas than expected. The AFA analysis across the 35 languages also revealed significantly more coronal codas ($M = 133.8, S.D. = 11.67$) than non-coronal codas ($M = 92.03, S.D. = 11.26$). This difference was significant, $t(34) 2.67, P < .01$, two-tailed. Twenty-seven out of the 35 individual languages had more words ending in coronal codas than other places of articulation. Of these 27 languages, 22 had significantly more coronals than other places of articulation based on individual chi-square goodness-of-fit tests. Note that with both the EFA and AFA analyses, individual languages are found to pattern differently, but the general tendency holds across the majority of the languages, where coronals are preferred.

Across languages, the preferred syllable shape has a sonorous ending; thus, the preferred syllable ends with a final vowel. When final consonants are present, it has been claimed that the preferred syllable-final consonants are sonorants, such as nasals and liquids (Clements, 1990: 303). This preference is found in typological surveys where languages tend to favor obstruents in initial position over final position (Coberly, 1985). There is also an unrestricted universal in languages, such that the presence of an obstruent coda implies the presence of a sonorant coda (Clements, 1990; Eckman and Iverson, 1994; Fonte, 1996). The preference for sonorant codas is based on more traditional observations and interpretations of universal grammar and markedness. This claim is supported by an examination of the 35 languages, which revealed that all languages with obstruent codas had sonorant codas as well.⁶

A chi-square goodness-of-fit test on the EFA revealed that there were more sonorant codas than expected by chance, $\chi^2(34, N = 7906) = 197.78, P < .001$. Furthermore, the same test performed on each language indicated that 14 languages had significantly more

⁵ In some languages dorsal is the unrestricted place feature in coda position (Rice, 1993). To determine whether a language's unmarked place feature is coronal or dorsal requires an analysis of a language's phonotactics and phonological processes. This is beyond the scope of the study. One can argue, however, that although both coronal and dorsal are unmarked in coda position, coronal is the least marked of the two because it is more frequently the unmarked place of articulation in coda position. For a similar argument, see Fonte (1996).

⁶ Notice that previous data and analyses have collapsed word-medial and word-final codas, whose distributions differ (Itô, 1986; Piggott, 2003); therefore, it was necessary to determine the sonority preferences based solely on word-final codas.

sonorant codas and six languages had significantly fewer sonorant codas than expected.⁷ An AFA determined whether there were significantly more sonorant than obstruent codas in languages' lexicons. Unlike with the EFA, no preference was found for languages to have either sonorants or obstruents. There were numerically more obstruent codas ($M = 132.8$, $S.D. = 12.03$) than sonorant codas ($M = 111.06$, $S.D. = 11.46$); however, this difference was not significant, $t(34) = 1.09$, n.s., two-tailed. Twenty-two languages had more words ending in obstruent codas than words ending in sonorant codas. Of these 22 languages, 15 had significantly more obstruents than sonorants based on individual chi-square goodness-of-fit tests. It is not surprising that the AFA did not reveal a significant difference between the number of sonorant versus obstruents in languages, given the nature of languages' inventories. Languages' phoneme inventories are largely composed of obstruents (voiced and voiceless stops and fricatives with three different places of articulation), whereas sonorants are restricted to nasals (usually just voiced nasals with three different places of articulation) and liquids (which are almost always restricted to coronal place of articulation). Because languages have a larger set of obstruent phonemes than sonorant phonemes, this accounts for why a greater number of obstruents than sonorants are found across languages, despite sonorants being universally preferred in this position.⁸ Importantly though, there was a higher proportion of sonorant codas across languages based on the EFA analysis, which illustrates that sonorants are more represented in the frequency of languages, than expected.

Based on the EFA and AFA results for coda place and sonority, and based on more traditional ways of determining markedness, the UGH should state that unmarked codas have coronal place of articulation and are sonorant. We are now in the position to operationalize the predictions of the UGH for English-learning children's early coda productions. The English coronal and sonorant codas are taken to characterize the cross-linguistic preferences, as this will allow for a direct comparison between the distribution of codas in English and child language data from English. In fairness to the UGH, it would be unlikely to find the affricates /tʃ/ and /dʒ/ among children's initial coda productions, despite the fact that they are coronals. These segments require sophisticated articulation skills and are therefore marked in terms of manner. This raises the question of how to fully determine the markedness of any particular segment given that some properties of a segment may be unmarked whereas other properties may be marked. Because of these complicating factors, it is not clear how an exact prediction of coda acquisition should be formulated. One needs to consider all the possible cross-linguistic characteristics of codas and how they interact, for example, the language-specific phoneme inventory and the language's phonological processes. Such an enterprise is beyond the scope of this study. However, even if one were able to take all of these factors into account, the result would be that the UGH would become less language specific. Despite these problems with

⁷ Arguments have been made for the special status of nasals in coda position (Fonte, 1996; Piggott, 2003). Thus, it is possible that the sonority preference was carried by languages' preferences for nasal codas. An examination of the sonority preference across the 35 languages revealed although 15 languages had more nasal codas than expected, there were 20 languages with more liquid codas than expected. Thus, the special status of nasals in coda position is not supported by these data.

⁸ This argument also holds for coronals because there are usually more coronals in a language's inventory than other places of articulation. However, the AFA analysis for place of articulation was significant, whereas the sonority analysis was not significant.

operationalizing a UG hypothesis, a reasonable starting point for the UG based hypotheses of coda acquisition (based on previous implementations of UG in acquisition) is that the UGH predicts that children's initial coda productions should bear coronal place of articulation over labial or dorsal place of articulation and will preferably be sonorant over obstruent.⁹ Still the UGH can exclude those sounds known to be late acquired or marked along other dimensions (1).¹⁰

- (1) Universal Grammar Hypothesis (UGH):
Children will produce coronal codas (t, d, s, z, n, l, r) more than labial or dorsal codas, and sonorant codas (m, n, l, ŋ, r) more than obstruent codas in word-final position.

3. Codas in the input to English-speaking children

There are a number of studies in the literature that have looked at the distribution of phonemes in English (e.g., Dewey, 1923; Roberts, 1965); however, no comprehensive study has examined the distribution of codas with respect to markedness, nor have these studies necessarily separated codas from onsets in their analyses. In addition, studies that have distinguished between onset and coda position fail to control for other prosodic positional effects, such as stress. For these reasons it was necessary to establish the distribution of word-final codas in English CVC words.¹¹

Data were taken from a corpus of Child Directed Speech (CDSC). To create this corpus, parental speech was taken from CHILDES (MacWhinney, 2000) studies involving children between the ages of 19–28 months. CVC words from the corpus were compiled and the number of word-final consonants for each of the English phonemes was recorded (see Zamuner, 2003 for a full description). All proper names were then excluded, given the assumption that this type of input varies greatly across children due to popular culture and different names. The resulting CVC word corpus consisted of 40,822 tokens, with 604 unique types. Phonetic transcriptions were obtained from the electronic version of the Webster's Dictionary of American English (Webster, 1964). The predictions here will be based on the token counts of the CDSC for the following reasons: this source provides the

⁹ One possibility is to evaluate children's phonological inventory and system as it develops over time. Once children have begun acquiring codas, it is possible that markedness is reflected in the type and order of contrasts as children expand their systems. These type of acquisition data are currently not available, but they would provide a promising further test case of the UGH versus the SLGH. Thanks to an anonymous reviewer for suggesting this analysis.

¹⁰ Although it would also be possible to collapse these hypotheses and argue that the unmarked coda is a coronal sonorant, e.g., /n/, one must also consider that the unmarked voiced feature in coda position is voiceless (Lombardi, 1991). Combining the three preferences would predict that the unmarked coda would be a voiceless, coronal sonorant, e.g., /ŋ/. However, the unmarked feature for sonorants is voiced (e.g., Steriade, 1995). Thus, the unmarked features in coda position are not compatible, and the hypotheses have not been combined to avoid an inaccurate characterization of cross-linguistic preferences in this position.

¹¹ Stoel-Gammon (Stoel-Gammon, 1995, cited in Stoel-Gammon, 1998) examined the phonological structure of children's early words from the MacArthur Communicative Development Inventory (CDI) (Fenson et al., 1993). An analysis of children's early words revealed that more target words had obstruents than sonorants, and that most word-final consonants were coronal.

closest approximation to children's input between the ages of 19 and 28 months, and token frequency has been shown to predict acquisition more accurately than type frequency (Marchman and Plunkett, 1989).

English codas were examined to see whether they reflect the universal preference for coronal place of articulation and sonority in coda position, which could confound analyses. Results of an EFA indicated that English has more coronal codas than expected ($\chi^2(1, N = 40822) = 5629.93, P < .001$). An AFA indicated that there are more coronal codas than other places of articulation ($\chi^2(1, N = 40822) = 14632.15, P < .001$). If children learning English initially produce coronal codas, it is not clear whether this fact reflects universal place of articulation preferences, or the distribution of place in the ambient language. Similarly, an EFA showed that English has more sonorant codas than expected ($\chi^2(1, N = 40822) = 2951.34, P < .001$). Importantly, an AFA revealed that English has significantly more obstruent than sonorant codas ($\chi^2(1, N = 40822) = 3552.25, P < .001$). That is, in terms of raw frequency, English has more obstruent codas, which is the opposite of what is found cross-linguistically based on traditional interpretations of cross-linguistic markedness, and based on the EFA analysis of sonority (recall that no significant effect was found for the AFA across languages). Thus, the analysis of input to English-speaking children reveals a mismatch between cross-linguistic and language-specific data of the sort mentioned earlier. Examining children's production of sonorants and obstruents in English then provides an interesting test case for the issues explored in this study. If children initially produce obstruents in coda position, we would have evidence that children produce the most frequent codas in their language, rather than the unmarked codas.

The input-based hypothesis for children's acquisition of codas based on the rank order of token coda counts of the SLGH is stated in (2).

(2) Specific Language Grammar Hypothesis (SLGH):

Children will produce the most frequent codas in word-final position: $t > r > n > d > z > k > s > l > m > v > j > g > p > \theta > \eta > tʃ > f > dʒ > b > ʒ, \delta$.

Note that the SLGH predicts that children will initially acquire the most frequent codas, not that the most frequent coda in the language will be acquired first, then the second most frequent coda and so on. Indeed, it would be difficult to argue that children would acquire the most frequent coda in the language, when there are slight variations in which coda is the absolute most frequent depending on the type of database one uses (Zamuner, 2003: 41). Note also that we represent the SLGH using a rank order. This is one of the problems that is faced in trying to tease apart the UGH and the SLGH. For the sake of consistency, we used rank order for the analyses. This is because the child language data were taken from a number of different sources, some of which only provided summary data with no raw scores included. Moreover, in order to compare the UGH and the SLGH in a consistent way, all analyses were based on rank orders.

Both the SLGH and the UGH predict that children will show an early preference for the codas: /t, n, r, d/. The SLGH differs from the UGH in that not all of the frequent codas in English are coronals or sonorants, for example /k/ is a frequent coda, yet it is an obstruent with dorsal place of articulation. Under the present characterization of UGH, the prediction is that children will initially produce coronal and sonorant codas. The interpretation of the SLGH offered here is that children will initially produce the most frequent codas, based on

the distribution of coda exemplars in the input. The characterization of UGH adopted here makes more general predictions, i.e., that the initial segment will be coronal/sonorant, but not a particular coronal/sonorant. The question then arises as to whether the more specific predictions of the SLGH can be abstracted to produce more general predictions so the two theories can be more directly comparable. One clear difference between the two approaches is that the SLGH predicts that frequency distributions will be random with respect to phonetic features like coronal and sonorant. In contrast, the UGH predicts that coronals as a group will be more frequent than non-coronals, and that sonorants as a group will be more frequent than obstruents. If adult input (English) and child data both show a random distribution of consonant features, and if the rank orders are significantly correlated, this will be evidence for the SLGH. In contrast, if the adult input and child language data show groupings of coronal and sonorant codas, this finding will be consistent with both the UGH and the SLGH. To address this issue, the distribution of English codas were analyzed for feature-based grouping.

The rank order of the 21 English codas based on the token word counts from the CDSC were analyzed for evidence of coronal, dorsal or labial place of articulation using the Kruskal–Wallis one-way analyses of variance (ranks). To compute the Kruskal–Wallis statistic, codas were ranked according to their frequency. The groups of coronal, dorsals and labial codas were not significantly represented in the rank order of English codas, $H(2, N = 21) = 1.22, P = .54$. That is, despite the fact that the average rank of coronal codas in English is higher than that of dorsal or labial codas (12 versus 11 and 8.4, respectively), this difference does not approach statistical significance. Similarly, a test was done to measure whether sonorants or obstruents are reflected in the rank order of English codas. The statistic used here was the Mann–Whitney U -test for two independent samples, which is similar to the Kruskal–Wallis one-way analysis of variance for three or more independent samples. Results revealed that sonority was marginally reflected in the rank order of English codas, $z = 1.45, P = .07$. Although the results were not statistically significant, the pattern was in the direction one would predict based on cross-linguistic preferences in this position. Compare the mean rank of sonorants (14.6) to the mean rank of obstruents (9.88).

In summary, the input to English-speaking children shows a mismatch with cross-linguistic data with respect to place of articulation, but not with respect to sonority. The UGH predicts that more children will produce coronal codas. In contrast, the SLGH predicts that children will produce codas that reflects (2) above, and will not show a significant place-based grouping of codas. The data for sonority are less clear and suggest evidence for sonority-based grouping cross-linguistically and in the input (i.e., in English). If child language data also show a grouping effect, then it will be difficult to distinguish the two hypotheses with respect to sonority. We now turn to child language data to test these hypotheses.

4. Codas in the productions of English-speaking children

Previous research looking at coda acquisition has revealed that children's early codas are often obstruents (Fikkert, 1994; Goad, 1997; Salidis and Johnson, 1997; Velten, 1943). More specifically, others have noted the special status in coda position of voiceless obstruents and

nasals over voiced obstruents (Bernhardt and Stemberger, 1998; Kehoe and Stoel-Gammon, 2001). The research here explores whether the same patterns hold across data from a larger group of children. Part of the data analyzed is taken from a number of previously published studies whose findings and/or primary data have been coalesced to provide as clear a picture of possible on aspects of children's coda productions. Data from previous studies are restricted to monolingual children acquiring English, under three-years of age. The other source analyzed was based on data from the Higgenson database in CHILDES Higginson, 1985), which provided data from April (1;10–2;11) and June (1;3–1;9). Analyses from these children productions were restricted to CVC monosyllabic words.

Previous research has shown mixed results with respect to the frequency of codas in the input and children productions. Bernstein-Ratner (1994) looked at the frequency of codas in the input with respect to children's coda productions, and no relationship was found. (No statistics were provided for the analyses.) In her study, children's final consonant productions at the preverbal, one-word stage and multi-word stage were compared to the maternal input. One limitation of this study is that the frequency of codas was collapsed across prosodic positions and word types. This can drastically change findings, given that the distribution of codas in English differs according to prosodic position and that children's early words are often only one syllable in length. In addition, Bernstein-Ratner's study used a phonetic analysis to convert the adult orthographic representations to phonetics, and from this analysis, the frequency of sounds in the adult input was calculated. Given this method, the adult input is calculated as having no instances of /z/, although final "s" is produced as /z/ (e.g., "cries", /kraɪz/). These reasons might account for why no relationship was found between the input and children's production of codas. Other work by (Stoel-Gammon, 1995, cited in Stoel-Gammon, 1998), has found various effects depending on the type of analysis. Her study tested whether the more frequent phonological features are acquired early and are less prone to errors in production. Results were mixed, and most strikingly, not supported for final fricatives which are frequent in the input, yet as a class, are acquired late. The difficulty in a feature analysis, however, is that although the class of fricatives might be acquired late, certain fricatives are acquired relatively early. Stoel-Gammon also correlated the frequency of codas based on the type counts of words from the MacArthur Communicative Development Inventory (CDI) (Fenson et al., 1993) to the accuracy of 3-year-olds coda productions from Templin (1957). At the phonemic level, a significant correlation was found between the frequency of occurrence in target words and the accuracy of children's productions. A recent study by Stites et al. (2004) also examined the role of frequency vs. markedness in the acquisition of codas, based on longitudinal data from two children acquiring English. They suggest that one child's data are more consistent with markedness effects, while the other child's data are more consistent with frequency. While a full comparison of differences in data analyses is beyond the scope of this paper, we can note briefly, that their analyses do not test for whether children produced the more frequent unmarked codas over less frequent unmarked codas. This is a crucial test. If the child who is argued to follow a markedness path also shows a preference for the more frequent unmarked codas, this would be a confound. Both a markedness and a frequency account would be compatible with the acquisition data.

Following Stoel-Gammon (1985), two primary methods were used to analyze children's productions. The first is referred to as an "independent analysis". This analysis measures the codas children produce, without consideration of the target sounds children attempt. This

“provides a complementary view of the child’s phonological system by describing the sounds [...] produced regardless of their relation to the adult model” (Stoel-Gammon, 1985: 505). For example, “coat” produced as [køk] would be coded as a coda production of /k/, with no evaluation of its incorrectness to the target /t/. Using this evaluation method, codas initially produced by children are given in Table 1 and the results are summarized in (3), which gives the most-to-least frequent codas initial produced by children. Given the variation inherent in looking at data across previously published research, the summary represents either the first codas produced by children (either stated by the author or discerned from the primary data) or the five most frequently produced codas (here not a single coda listed or determined as produced first, therefore, all codas produced by the child/children are included). Based on data from 41 children, the coda that children are most likely to produce is /t/.

- (3) Child language patterns for English codas established by independent analysis:
t > n > k > d, m > ? > p > tʃ, l > b, f, F, η, r, s.

The second method of analysis is a “relational analysis”, “which compares the child’s pronunciation of a word with the adult standard form. The comparison may concentrate on the correct sound produced and describe the emergence and mastery of phonemes of the adult language” (Stoel-Gammon, 1985: 505). Thus, this method examines the accuracy of children’s coda productions. The same example of “coat” produced as [køk], is coded as an error in the child’s production of /t/, and would not be counted in a relational analysis. A summary of the data used for the relational analysis is given in Table 1, and results using this evaluation method are given in (4), in order of the most frequent codas to the least frequent codas produced correctly. Results are based on data from 28 children. Note that in both the independent and relational analyses, x > y indicates either that more children produced coda x as their first coda, or that more children produced coda x than coda y.

- (4) Child language patterns for English codas established by relational analysis:
k > t > n > d, s > b, f, l, m, η, p, r, z.

One underlying assumption is that children are able to correctly perceive codas. Although this is often assumed in child language phonological research, the results from this summary would be strengthened with positive evidence that demonstrates children were able to correctly perceive codas. There are a few studies that support the claim that children’s perception is adult-like. Studies in infant perception research have shown that infants aged seven and a half months are able to perceive the difference between minimal pairs that differ only on the codas’ place of articulation (Tincoff and Jusczyk, 1996). Older children, between the ages of three and four-years-old age also show the same ability (Edwards et al., 2002). Children at this age perform above chance in discriminating between minimal pairs that differ in the codas’ place of articulation, e.g., ‘cap’ versus ‘cat’. Given the results from these types of studies, we argue that children’s perception was adult-like and that the summaries drawn from these data are accurate.

5. Evaluation of the UGH and the SLGH

We are now in the position to evaluate the UGH and the SLGH with respect to data from children acquiring English. The UGH for coda acquisition was established according to

Table 1
Results from independent analysis and relational analysis of children's coda productions

References	Subject(s)	Independent analysis first or five most produced codas	Relational analysis first or five most produced codas
Bernstein-Ratner (1994)	Nine children	n > t > k > d > m (13–25M)	n/a
Braine (1974)	Jonathan	? (19M)	n/a
French (1989)	Andrew	tʃ (18M)	s (19M)
Higginson (1985)	April	t > n > s > k > p (22–23M)	b (22–35M)
Higginson (1985)	June	t > k > l > p > n (15–21M)	f (15–21M)
Holmes (1927)	Mollie	k, t, d (18M)	k, t, d (13M)
Ingram (1974)	Jennika	t (15M)	t (15M)
Massar (unpublished data)	Ben	t (13M)	t (13M)
Menn (1971)	Daniel	r, F (unspecified fricative), f, b ~ m (22M)	r (22M)
Ohala (1992)	Joey	n, l, ? (18M)	n, l (18M)
Olmsted (1971)	17 children	n/a	1/2 subjects correct – k ^a (15–23M)
Salidis and Johnson (1997)	Kyle	n/a	d, k, m, p, s, t, z (11M)
Shibamoto and Olmsted (1978)	“F”	t ^h ~ t (14M)	n/a
Stoel-Gammon (1985) ^b	19 children	t (18M)	n/a
Taken from Vihman (1996)	Alice	? (14M)	n (14M)
Taken from Vihman (1996)	Molly	ŋ, n, ʔ, t, tʃ (13M)	ŋ, n, t (11M)
Taken from Vihman (1996)	Timmy	p (11M)	n/a

^a Correct is defined as at least 50% of the phoneme's productions were produced correctly.

^b A sound was considered part of the child's inventory if produced in at least two unique words.

cross-linguistic patterns, which are taken to reflect UG. Thus, if language acquisition is mediated by innate properties of UG, the prediction is that children will prefer coronal and sonorant codas. The UGH for coda acquisition is restated in (5).

- (5) Universal Grammar Hypothesis (UGH):
Children will produce coronal codas (t, d, s, z, n, l, r) more than labial or dorsal codas, and sonorant codas (m, n, ŋ, l, r) more than obstruent codas in word-final position.

The predictions of the SLGH for coda acquisition based on the rank order of codas, from the token word counts of the CDSC, and restated in (6).

- (6) Specific Language Grammar Hypothesis (SLGH):
Children will produce the most frequent codas in word-final position: t > r > n > d > z > k > s > l > m > v > ʃ > g > p > θ > ŋ > tʃ > f > ɸ > b > ʒ, ð.

The UGH and the SLGH have different predictions for children's preferences for codas with coronal place of articulation, in that the SLGH predicts children will also initially produce the frequent coda /k/, despite it having dorsal place of articulation. With respect to sonority, the UGH makes a strong prediction that children will display a preference for sonorant codas. The SLGH predicts a preference for sonorants (recall that sonorants were marginally reflected in the rank order of English codas), but not as clearly as with the UGH.

Children's productions of codas in English were characterized in two ways. Data were coalesced from previous published studies and data from CHILDES. These data established the codas children initially produce (independent analysis) and the codas children initially produce accurately (relational analysis). The results are summarized in Table 2. There is both consistency and inconsistency found across the different analyses. Take for example the five most frequently produced codas as determined by the two analyses. Both results include the codas /t, n, k, d/; however, the independent analysis contains /m/ and the relational analysis contains /s/. Both analyses were analyzed separately with respect to the hypotheses.

To evaluate the UGH, children's coda productions were measured for whether coronal codas and sonorant codas were the most often produced codas by children acquiring English. Based on the Kruskal–Wallis one-way analysis of variance, results showed no evidence that the average rank of coronal codas was significantly higher or distinct from that of dorsal and labial codas: independent analysis ($H(2, N=15) = 0.35, P = .84$), relational analysis ($H(2, N=13) = 2.45, P = .29$). Similarly, the Mann–Whitney U -test for two independent samples was used to determine whether children's coda productions were more frequently sonorants or obstruents. Results showed no evidence for sonorants and obstruents forming distinct groups within the rank order of English codas produced by children: independent analysis ($z = 0, P = .5$), relational analysis ($z = 0.81, P = .21$).

Table 2
Child language patterns for English

Analysis	Child language patterns: English
Independent analysis	t > n > k > d, m > ? > p > tʃ, l > b, f, F, ŋ, r, s
Relational analysis	k > t > n > d, s > b, f, l, m, ŋ, p, r, z

Although coronals and sonorants are among the set of most frequently produced codas by children (see Table 2), there are also other codas produced that are not predicted by the UGH, namely /k/ (and in one case /m/). The production of /k/ in coda position cannot be accounted for under the current interpretation of UG; however, a frequency analysis can offer an explanation for this pattern — that is, /k/ is the sixth most frequent coda in the input. Children's coda productions established through two different means were significantly correlated to the rank order of codas based on the frequency of codas: independent analysis ($r_s = .51$, $n = 13$, $P < .05$), relational analysis ($r_s = .59$, $n = 11$, $P < .05$), one-tailed. (Another analysis was also done in which children's productions were correlated to the raw frequency of codas in English. Both correlations were significant.)

One possibility for why the SLGH appears to better account for the acquisition data than the UGH is that the analyses of the UGH might be more stringent. A analogous analysis of the SLGH was done where frequency was treated as a category. Recall that the SLGH predicted that children will prefer coronal codas (13 in English) over codas with dorsal and labial (8 in English) place of articulation. The SLGH was then tested by determining whether children preferred the 13 most frequent codas in their productions over the 8 less frequent codas. Results of the Mann–Whitney *U*-test showed significant effects in one operationalization of child language data and a marginal effect in the second: independent analysis ($z = 1.85$, $P < .05$), relational analysis ($z = 1.18$, $P = .12$). The UGH was also tested by determining whether children show evidence for the groupings of sonorants (5 in English) and obstruents (16 in English). The equivalent test with the SLGH is whether there children prefer the 5 most frequent codas over the 16 less frequent codas in English. Results of the Mann–Whitney *U*-Test also showed marginal significance across the two operationalizations of child language data: independent analysis ($z = 1.39$, $P = .08$), relational analysis ($z = 0.88$, $P = .19$). Therefore, even though the UGH might be more stringent, it is unlikely that this can account for its failure to accurately predict children's acquisition of codas.

6. Conclusion

The starting point of this research was the observation that child language reflects cross-linguistic markedness. The traditional account of this observation is that it reflects Universal Grammar. UG (established primarily through cross-linguistic research) has been argued to constitute the initial hypotheses children entertain for language. Thus, if the universally unmarked syllable shape is a CV syllable, children should exhibit an initial sensitivity to these syllable types. It was noted, however, that one must consider the fact that cross-linguistic markedness is often at least partially mirrored in a single language's distribution of sounds. This fact brings into question the relationship between child language and UG; child language patterns can also be characterized as reflecting the distribution of sounds in the ambient language. Thus, although one could argue that children initially produce unmarked structures due to innate UG, one could equally argue that children initially produce unmarked structures because these patterns are the most frequent in the language to which they are exposed. The argument here is not that language is acquired based on the input alone, but rather that both theories need to be considered as they often make the same predictions. The points where cross-linguistic and language-

specific data mismatch, provide test cases for theories of acquisition. Child language data can then determine whether acquisition is more consistent with innate theories of acquisition (UGH) or with a less constrained acquisition mechanism (SLGH).

To evaluate child language with respect to the UGH, children's coda productions were analyzed to determine whether the preferred codas had coronal place of articulation and whether they were sonorant codas. Results showed no preference for these codas in children's productions, thereby, refuting the UGH. It is an open question whether other interpretations of UG can capture these data, taking into consideration the various features that segments can bear and the markedness of such features, such as voice, continuancy, sonority, and place of articulation. The measure of evaluating or determining which features or which feature combinations are marked, however, it is not clear. It is important to note that this study shows that the most common conception of UG is not adequate in explaining children's early coda productions. It is worth stating again that the characterization and evaluation of UG offered here is not exhaustive, but rather it is taken as a starting point for further research. UG is often given as an account for the patterns seen in children's early language productions and as a means to explain how children's phonological systems develop. The work here illustrates the complexity faced in predicting specific acquisition patterns from this model and in applying this model to existing data from child language.

To evaluate the SLGH, analyses determined whether there were significant correlations between the codas produced by children and the frequency of those codas in English. Across both operationalizations of child language data, there was a significant relationship between children's coda productions and the frequency of codas in English. Thus, as formalized in this study, the results favor the SLGH over the UGH as an account of children's acquisition of codas in CVC words. In this sense, children organize and build their phonological knowledge of word-final consonants based on frequently occurring patterns in the ambient language. Children do not show evidence for the universally preferred codas in their productions, and this finding suggests that children do not necessarily come to the acquisition task with pre-specified knowledge of what are the optimal codas.¹² One must also consider the source of cross-linguistic markedness. Cross-linguistic patterns are also reflected in the distribution of a single language's sound patterns, and this suggests that these patterns reflect the basic human abilities of perception and/or production, rather than innate linguistic knowledge.

¹² The results also present an interesting finding for the ways in which children might calculate frequency. English data were analyzed with respect to the universal preferences of coronal place of articulation and sonority in coda position. Both EFA and AFA analyses showed evidence for the favored status of coronal place of articulation. Results of sonority revealed that English has more sonorant codas than expected (EFA), and a greater number of obstruent codas (AFA). This latter finding can be accounted for by appealing to English's coda inventory, which has more obstruents (16) than sonorants (5). It has been argued in this study that children do not show a preference for the unmarked codas which are sonorant, thus refuting the UGH. This further suggests, however, that children do not attend to proportions of obstruents and sonorants in the input (EFA), but rather, that children attend to input frequencies based on the raw number of occurrences of codas in the language (AFA). Children also do not appear to attend to the rank order of segment types. Recall that sonority was marginally reflected in the rank order of English codas, but no preference for sonority was found in the rank order of children's coda productions.

Previous research looking at coda acquisition has noted that obstruents are acquired before sonorants in coda position (Fikkert, 1994; Goad, 1997; Salidis and Johnson, 1997) and that voiceless obstruents and nasals are preferred over voiced obstruents in coda position (Bernhardt and Stemberger, 1998; Kehoe and Stoel-Gammon, 2001). Accounts for this have argued that children produce the maximal contrast in coda position from the preceding vowel (Fikkert, 1994), that this reflects the structure of these segments in which they are syllabified as the onsets of empty headed syllables (Goad, 1997), or that this reflects the default coda (margin) features and articulatory and perceptual ease (Bernhardt and Stemberger, 1998; Kehoe and Stoel-Gammon, 2001). A closer look at the patterns from previous research (and noted in previous research) reveals that children's production of codas is often varied, and children produce a variety of segments as their initial codas. Thus, looking at acquisition in terms of classes of segments is an abstraction from the data which might inhibit the developmental picture. When we look at the words and segments that children are mostly like to acquire and produce as their first words, an alternative account for the data can be found. Here the patterns in the ambient language drive the acquisition effects. Yet, any theory of acquisition that relies on the input to organize children's linguistic representations needs to ultimately account for how and why frequently occurring properties are attended to by children, and how children ultimately move beyond what is frequent to ultimately acquire a full linguistic system. Demonstrating that children are sensitive to frequency information does not explain how language is acquired by children. In this sense, input based models of acquisition are plagued with the same problems found in UG and nativist based accounts of acquisition, where demonstrating that children have innate linguistic knowledge still does not explain the language acquisition process.

Although input patterns are argued to account for patterns seen in child language acquisition, it is possible that the findings presented in this study reflect a universal "ease of articulation" or "ease of perception" in coda position. A confound arises though because "ease" can vary according to language experience, such that frequent sounds might be less difficult than infrequent sounds (Bernhardt and Stemberger, 1998). This is further illustrated with research that shows that phoneme acquisition can vary across languages depending on the relative frequency of phonemes (e.g., Macken, 1995; Pye et al., 1987; Ingram, 1988). These results may reflect the sounds of the language that children are most likely to attempt. It is difficult to evaluate an "ease of articulation" or "ease of perception" theory because there is no scale or definition for ease of articulation or perception. However, a number of facts from child language suggest that these alternative accounts cannot solely account for the results presented in this study. First, sensitivity to patterns in the input is demonstrated by infants, even before they begin producing meaning speech. Second, children's imitated productions are often phonologically advanced from their spontaneous productions. Thus, it appears that children are not limited by articulatory or perceptual constraints, but rather that children's errors are largely influenced by their ability to access stored representations.

In summary, the long noted relation between cross-linguistic patterns and child language data has been traditionally taken to support a view of language acquisition in which children's first productions reflect a biologically given Universal Grammar. However, the sound patterns that are found across languages are often frequent in a particular language, suggesting that children's utterances may reflect their own linguistic input. A comparison of codas across languages, in adult English and in English-speaking children

revealed that the child language was more similar to the input than to cross-linguistic data. These data suggest that the relation of cross-linguistic patterns and child language data is different from and more complex than the traditional UG interpretation.

Appendix A

Languages used in cross-linguistic analyses organized by language (classifications taken from Ethnologue, Grimes and Grimes, 1999).

Language	Language family	References
Alawa	Australian	Sharpe (1972)
Arabic	Afro-Asiatic	Abu-Absi (1995)
Armenian	Indo-European	Kozintseva (1998)
Brahui	Dravidian	Bray (1986)
Cantonese	Sino-Tibetan	Huang (1970)
Daga	Trans-New Guinea	Murane (1974)
Daur	Altaic	Wu (1998)
Dehong	Daic	Luo (1998)
Dehu	Austronesian	Tryon (1967a,b)
Doyayo	Niger-Congo	Wiering and Wiering (1994)
Dutch	Indo-European	Baayen et al. (1993)
Hebrew	Afro-Asiatic	Doniach and Kahane (1996)
Jacalteco	Mayan	Craig (1977)
Karachay	Altaic	Seegmiller (1996)
Khmer	Austro-Asiatic	Headley et al. (1977)
Kurdish	Indo-European	McCarus (1967)
Latvian	Indo-European	Fennell and Gelsen (1980), Nau (1998) and Turkina (1982)
Lisu	Sino-Tibetan	Bradley (1994)
Louisiana Creole French	Creole	Valdman et al. (1988)
Luo	Nilo-Saharan	Stafford (1967)
Mam	Mayan	England (1983)
Mbay	Nilo-Saharan	Keegan (1997) and Keegan et al. (1996)
Munari	Austro-Asiatic	Sinha (1975)
Nivkh (Gilyak)	Language Isolate	Gruzdeva (1998)
Pima Bajo	Uto-Aztecan	Estrada Fernández (1996)
Pulaar	Niger-Congo	Niang (1997)
Rhade	Austronesian	Tharp and Y-Bham (1980)
Salt-Yui	Trans-New Guinea	Irwin (1974)
Serbian	Indo-European	Benson (1982)
Telefol	Trans-New Guinea	Healey (1977) and Healey and Healey (1977)
Thai	Daic	Manich Jumsai (1958)
Tibetan	Sino-Tibetan	Goldstein and Normang (1970)
Totonac	Totonacan	MacKay (1999)
Turkmen	Altaic	Garrett et al. (1996)
Ukrainian	Indo-European	Danylenko and Vakulenko (1994)

References

- Abu-Absi, S., 1995. Chadian Arabic, Lincom Europa, Munich, Germany.
- Baayen, R.H., Piepenbrock, R., van Rijn, H., 1993. The CELEX lexical database [CD-ROM]. Linguistics Data Consortium, University of Pennsylvania.

- Benson, M., 1982. Standard English-SerboCroatian, Serbo-Croatian-English Dictionary, Cambridge University Press, New York.
- Bernhardt, B.H., Stemberger, J.P., 1998. Handbook of Phonological Development: From the Perspective of Constraint-Based Nonlinear Phonology, Academic Press, San Diego, CA.
- Bernstein-Ratner, N., 1994. Phonological analysis of child speech. In: Sokolov, J.L., Snow, C.E. (Eds.), Handbook of Research in Language Development Using Childes. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 324–372.
- Blevins, J., 1995. The syllable in phonological theory. In: Goldsmith, J. (Ed.), The Handbook of Phonological Theory. Blackwell, Oxford, pp. 206–244.
- Blumstein, S., 1973. A Phonological Investigation of Aphasic Speech, Mouton, The Hague.
- Booij, G., Rubach, J., 1984. Morphological and prosodic domains in lexical phonology. *Phonology Yearbook* 1, 1–27.
- Borowsky, T., 1986. Topics in the lexical phonology of English. Unpublished doctoral dissertation, University of Massachusetts at Amherst.
- Bradley, D., 1994. A Dictionary of the Northern Dialect of Lisu, Australian National University, Canberra, Australia.
- Braine, M.D.S., 1974. On what might constitute learnable phonology. *Language* 50, 270–299.
- Bray, D., 1986. The Brahui Language, Gian Publishing House, Delhi, India.
- Chomsky, N., 1981. Lectures on Government and Binding, Foris, Dordrecht, The Netherlands.
- Clements, G.H., 1990. The role of sonority cycle in core syllabification, In: Kingston, J., Beckman, M. (Eds.), 1990. Papers in Laboratory Phonology: Between the Grammar and Physics of Speech, vol. I. Cambridge University Press, Cambridge (pp. 283–333).
- Clements, G.N., Keyser, S.J., 1983. CV Phonology: A Generative Theory of the Syllable, MIT Press, Cambridge, MA.
- Coberly, M.S., 1985. Initial and final consonant preferences across languages, in babbling, and in children's early speech. Unpublished doctoral dissertation, University of Colorado.
- Craig, C.G., 1977. The Structure of Jacalteco, University of Texas Press, Austin, TX.
- Danylenko, A., Vakulenko, S., 1994. Ukrainian, Lincom Europa, Munich, Germany.
- Demuth, K., 2001. Prosodic constraints on morphological development, In: Weissenborn, J., Höhle, B. (Eds.), 2001. Approaches to Bootstrapping: Phonological, Syntactic, and Neurophysiological Aspects of Early Language Acquisition, vol. 2. John Benjamins, Amsterdam (pp. 2–21).
- Demuth, K., Fee, E.J., 1995. Minimal prosodic words in early phonological development. Unpublished manuscript, Brown University and Dalhousie University.
- Derwing, B.L., 1992. A Pause-Break task for eliciting syllable boundary judgments from literate and illiterate speakers: preliminary results from five diverse languages. *Language and Speech* 35, 219–235.
- Dewey, G., 1923. Relative Frequency of English Speech Sounds, Harvard University Press, Cambridge, MA.
- Doniach, N.S., Kahane, A., 1996. The Oxford English–Hebrew Dictionary, Oxford University Press, New York.
- Eckman, F.R., Iverson, G.K., 1994. Pronunciation difficulties in ESL: coda consonants in English interlanguage. In: Yavas, T. (Ed.), First and Second Language Phonology. Singular Publishing Group, San Diego, CA, pp. 251–266.
- Edwards, J., Fox, R.A., Rogers, C.L., 2002. Final consonant discrimination in children: effects of phonotactic disorder, vocabulary size, and articulatory accuracy. *Journal of Speech, Language, and Hearing Research* 45, 231–242.
- England, N.C., 1983. A Grammar of Mam, a Mayan Language, University of Texas Press, Austin, TX.
- Estrada Fernández, Z., 1996. Pima Bajo, Lincom Europa, Munich, Germany.
- Faingold, E.D., 1990. The acquisition of syllabic and word structure: individual differences and universal constraints. *Language Sciences* 12, 101–113.
- Fallows, D., 1981. Experimental evidence for English syllabification and syllable structure. *Journal of Linguistics* 17, 309–317.
- Fennell, T.G., Gelsen, H., 1980. A Grammar of Modern Latvian, vol. 3. Mouton, The Hague.
- Fenson, L., Dale, P., Reznick, S., Thal, D., Bates, E., Hartung, J., Pethick, S., Reilly, J., 1993. MacArthur Communicative Development Inventories, Singular Publishing Group Inc., San Diego, CA.

- Ferguson, C.A., 1963. Assumptions about nasals: a sample study in phonological universals. In: Greenberg, J. (Ed.), *Universals of Language*. MIT Press, Cambridge, MA, pp. 53–60.
- Fikkert, P., 1994. *On the Acquisition of Prosodic Structure*. Holland Institute of Generative Linguistics, Leiden, The Netherlands.
- Fikkert, P., Freitas, M.J., 1997. Acquisition of syllable structure constraints: evidence from Dutch and Portuguese. In: Sorace, A., Heycock, C., Shillcock, R. (Eds.), *Proceedings of GALA '97 on Language Acquisition: Knowledge Representation and Processing*. Edinburgh University Press, Edinburgh, pp. 217–222.
- Fonte, I., 1996. *Restrictions on Codas: An Optimality Theoretic Account of Phonotactics*. Unpublished masters thesis, McGill University, Montreal, Canada.
- Franke, C., 1912. Über die erste Lautstufe der Kinder (C. Murchison and S. Langer, Trans.). In: Bar-Adon, A., Leopold, W.F. (Eds.), *Child Language: A Book of Readings*. Prentice-Hall, Englewood Cliffs, NJ, pp. 13–17.
- French, A., 1989. The systematic acquisition of word forms by a child during the first-fifty word stage. *Journal of Child Language* 16, 69–90.
- Garrett, J., Lastowka, G., Naahielua, K., Pallipamu, M., 1996. *Turkmen-English dictionary* [on-line]. Internet: <http://www.chaihana.com/dict.doc>.
- Gennari, S., Demuth, K., 1997. Syllable omission in the acquisition of Spanish. In: Hughes, E., Hughes, M., Greenhill, A. (Eds.), *Proceedings of the Annual Boston University Conference on Language Development*, vol. 1. Cascadilla Press, Somerville, MA (pp. 182–193).
- Goad, H., 1997. Codas, word minimality, and empty-headed syllables. In: Clark, E.V. (Ed.), *Proceedings of the Twenty-eighth Annual Child Language Research Forum*. Center for the Study of Language and Information, Stanford, CA, pp. 113–122.
- Goad, H., Brannen, K., in press. Syllabification at the right edge of words: Parallels between child and adult grammars. In: van der Weijer, J. (Ed.), *Issues in the Phonology–Phonetics Interface*. John Benjamins, Amsterdam, in press.
- Goldstein, M.C., Norngang, N., 1970. *Modern Spoken Tibetan: Lhasa Dialect*. University of Washington Press, Seattle, WA.
- Greenberg, J.H., 1954. A quantitative approach to the morphological typology of language. In: Spencer, R.F. (Ed.), *Method and Perspective in Anthropology*. University of Minnesota Press, Minneapolis, MN, pp. 192–220.
- Greenberg, J.H., 1963. Some universals of grammar with particular reference to the order of meaningful elements. In: Greenberg, J.H. (Ed.), *Universals of Language*. MIT Press, Cambridge, MA, pp. 58–90.
- Greenberg, J.H., 1966. *Language Universals, with Specific Reference to Feature Hierarchies*. Mouton, The Hague.
- Greenberg, J.H., 1974. *Language Typology: A Historical and Analytic Overview*. Mouton, The Hague.
- Greenberg, J.H., 1978. Some generalizations concerning initial and final consonant clusters. In: Greenberg, J.H. (Ed.), *Universals of human language*, vol. 2. Stanford University Press, Stanford, CA (pp. 243–280).
- Grimes, J.E., Grimes, B.F., 1999. *Ethnologue: Languages of the world* [CD-ROM]. Summer Institute of Linguistics, Dallas, TX.
- Gruzdeva, E., 1998. *Nivkh*. Lincom Europa, Munich, Germany.
- Halle, M., Vergnaud, J.-R., 1980. Three dimensional phonology. *Journal of Linguistics Research* 1, 83–105.
- Hammond, M., 1997. Vowel quantity and syllabification in English. *Language* 73, 1–17.
- Hammond, M., 1999. *The Phonology of English*. Oxford University Press, New York.
- Hammond, M., Moravcsik, E.A., Wirth, J.R., 1988. Language typology and linguistic explanation. In: Hammond, M., Moravcsik, E.A., Wirth, J.R. (Eds.), *Studies in Syntactic Typology*, vol. 17. John Benjamins, Amsterdam (pp. 1–22).
- Harris, J., 1997. Licensing inheritance: an integrated theory of neutralisation. *Phonology* 7, 315–370.
- Healey, A., 1977. *Telefol Phonology*. Australian National University, Canberra, Australia.
- Healey, P.M., Healey, A., 1977. *Telefol Dictionary*. Australian National University, Canberra, Australia.
- Headley Jr., R.K., Chhor, K., Lim, L.K., Cheang, L.H., Chun, C., 1977. *Cambodian–English Dictionary*, vol. 1–2. Catholic University of American Press, Washington, DC.
- Higginson, R.P., 1985. *Fixing-assimilation in language acquisition*. Unpublished doctoral dissertation, Washington State University.

- Holmes, U.T., 1927. The phonology of an English-speaking child. *American Speech* 2, 212–225.
- Huang, P.P., 1970. *Cantonese Dictionary: Cantonese–English, English–Cantonese*. Yale University Press, New Haven, CT.
- Hyams, N.M., 1986. *Language Acquisition and the Theory of Parameters*. Reidel, Dordrecht, The Netherlands.
- Ingram, D., 1974. Phonological rules in young children. *Journal of Child Language* 1, 49–64.
- Ingram, D., 1988. The acquisition of word-initial /v/. *Language and Speech* 31, 77–85.
- Irwin, B., 1974. *Salt-Yui Grammar*. Australian National University, Canberra, Australia.
- Itô, J., 1986. *Syllable Theory in Prosodic Phonology*. Garland Press, New York.
- Jakobson, R., 1968. *Child Language, Aphasia, and Phonological Universals*. Mouton, The Hague (original work published in 1941).
- Kahn, D., 1976. *Syllable-based Generalizations in English Phonology*. Indiana University Linguistics Club, Bloomington, IN.
- Kaye, J., Lowenstamm, J., Vergnaud, J.-R., 1990. Constituent structure and government in phonology. *Phonology* 7, 193–231.
- Keegan, J.M., 1997. *A Reference Grammar of Mbay*. Lincom Europa, Munich, Germany.
- Keegan, J.M., Marcel, N., Bertin, M.T., 1996. *Dictionary of Mbay*. Lincom Europa, Munich, Germany.
- Kehoe, M.M., Stoel-Gammon, C., 2001. Development of syllable structure in English-speaking children with particular reference to rhymes. *Journal of Child Language* 28, 393–433.
- Kirk, C., Demuth, K., 2004. Contextual effects on the acquisition of coda consonants: prosodic environment, position within the word, and word length. Unpublished manuscript, Brown University.
- Kozintseva, N.A., 1998. *Modern Eastern Armenian*. Lincom Europa, Munich, Germany.
- Levelt, C., Schiller, N., Levelt, W., 2000. The acquisition of syllable types. *Language Acquisition* 8, 237–264.
- Lombardi, L., 1991. *Laryngeal Features and Laryngeal Neutralization*. Garland Press, New York.
- Luo, Y., 1998. *A Dictionary of Dehong, Southwest China*. Pacific Linguistics, Canberra, Australia.
- MacKay, C.J., 1999. *A Grammar of Misanla Totonac*. University of Utah Press, Salt Lake City, UT.
- Macken, M., 1995. Phonological acquisition. In: Goldsmith, J. (Ed.), *The Handbook of Phonological Theory*. Blackwell, Oxford, pp. 671–696.
- MacWhinney, B., 2000. *The CHILDES Project: Tools for Analyzing Talk*, third ed. Lawrence Erlbaum Associates, Mahwah, NJ.
- Maddieson, I., 1984. *Patterns of Sounds*. Cambridge University Press, Cambridge, MA.
- Manich Jumsai, M.L., 1958. *Thai-English dictionary*, 48th ed. Chalermnit, Bangkok, Thailand.
- Marchman, V., Plunkett, K., 1989. Token frequency and phonological predictability in a pattern association network: implications for child language acquisition. In: Olson, G., Smith, E. (Eds.), *Proceedings of the Eleventh Annual Meeting of the Cognitive Science Society*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 179–187.
- McCarthy, J., 1979. *Formal Problems in Semitic Morphology and Phonology*. Garland Press, New York.
- McCarus, E.N., 1967. *Kurdish–English Dictionary, Dialect of Sulaimania*. Iraq University of Michigan Press, Ann Arbor, MI.
- Menn, L., 1971. Phonotactic rules in beginning speech. *Lingua* 26, 225–251.
- Mowrer, D., Burger, S., 1991. A comparative analysis of phonological acquisition of consonants in the speech of 2.5–6-year-old Xhosa- and English-speaking children. *Clinical Linguistics and Phonetics* 5, 139–164.
- Murane, E., 1974. *Daga grammar, from Morpheme to Discourse*. Summer Institute of Linguistics of the University of Oklahoma, Norman, OK.
- Nau, N., 1998. *Latvian*. Lincom Europa, Munich, Germany.
- Niang, M.O., 1997. *Constraints on Pulaar Phonology*. University Press of America, London.
- Ohala, D., 1992. A unified theory of final consonant deletion in early child speech. In: Gerfen, C., Pinar, P. (Eds.), *Coyote Papers: Working Papers in Linguistics from A → Z*. University of Arizona, Tucson, AZ, pp. 96–117.
- Ohala, D., 1996. *Cluster reduction and constraints in acquisition*. Unpublished doctoral dissertation, University of Arizona, Tucson, AZ.
- Olmsted, D.L., 1971. *Out of the Mouths of Babes: Earliest Stages in Language Learning*. Mouton, The Hague.
- Paradis, C., Prunet, J., 1991. *The Special Status of Coronals: Internal and External Evidence: Phonology and Phonetics* 2. Academic Press, San Diego, CA.

- Piggott, G.L., 1991. Apocope and the licensing of empty-headed syllables. *The Linguistic Review* 8, 287–318.
- Piggott, G.L., 1999. At the right edge of words. *The Linguistic Review* 16, 143–185.
- Piggott, G.L., 2003. The phonotactics of a ‘Prince’ language: a case study. In: Ploch, S. (Ed.), *Living on the Edge: 28 Papers in Honour of Jonathan Kaye*. Mouton, Berlin, pp. 401–405.
- Poole, I., 1934. Genetic development of articulation of consonant sounds in speech. *Elementary English Review* 11, 159–161.
- Pye, C., Ingram, D., List, H., 1987. A comparison of initial consonant acquisition in English and Quiché. In: Nelson, K.E., Van Kleeck, A. (Eds.), 1987. *Children’s Language*, vol. 6. Lawrence Erlbaum Associates, Hillsdale, NJ (pp. 175–190).
- Rice, K., 1993. Default variability: the coronal–velar relationship. *Toronto Working Papers in Linguistics* 12, 97–130.
- Roark, B., Demuth, K., 2000. Prosodic constraints and the learner’s environment: a corpus study. In: Howell, S.C., Fish, S.A., Keith-Lucas, T. (Eds.), 2000. *Proceedings of the 24th Boston University Conference on Language Development*, vol. 2. Cascadilla Press, Somerville, MA (pp. 597–608).
- Roberts, A.H., 1965. *A Statistical Linguistic Analysis of American English*, Mouton, The Hague.
- Salidis, J., Johnson, J.S., 1997. The production of minimal words: a longitudinal case study of phonological development. *Language Acquisition* 6, 1–36.
- Seegmiller, S., 1996. Karachay, Lincom Europa, Munich, Germany.
- Selkirk, E., 1982. Syllables. In: Hulst, H., van der Smith, N. (Eds.), *The Structure of Phonological Representations*. Foris, Dordrecht, The Netherlands, pp. 337–383.
- Sharpe, M.C., 1972. *Alawa Phonology and Grammar*, Australian Institute of Aboriginal Studies, Canberra, Australia.
- Shibamoto, J.S., Olmsted, D.L., 1978. Lexical and syllabic patterns in phonological acquisition. *Journal of Child Language* 5, 417–446.
- Sinha, N.K., 1975. *Mundari Grammar*, Central Institute of Indian Languages, Mysore, India.
- Spencer, A., 1996. *Phonology*, Blackwell, Cambridge, MA.
- Stafford, R.L., 1967. *An Elementary Luo Grammar, with Vocabularies*, Oxford University Press, London.
- Steriade, D., 1995. Underspecification and markedness. In: Goldsmith, J. (Ed.), *Handbook of Phonological Theory*. Blackwell, Oxford, pp. 114–174.
- Stites, J., Demuth, K., Kirk, C., 2004. In: Brugos, A., Micciulla, L., Smith, C.E. (Eds.), 2004. *Proceedings of the 28rd Boston University Conference on Language Development*, vol. 2. Cascadilla Press, Somerville, MA (pp. 564–567).
- Stoel-Gammon, C., 1985. Phonetic inventories, 15–24 months: a longitudinal study. *Journal of Speech and Hearing Research* 28, 505–512.
- Stoel-Gammon, C., 1995. What children say: A phonological analysis of the target words acquired by young American children. Paper presented at Child Language Seminar, Bristol University, UK, April.
- Stoel-Gammon, C., 1998. Sounds and words in early language acquisition: the relationship between lexical and phonological development. In: Paul, R. (Ed.), *Exploring the Speech–Language Connection*. Paul H. Brookes Publishing Co., Baltimore, MD, pp. 25–52.
- Templin, M., 1957. *Certain Language Skills in Children*, University of Minnesota Press, Minneapolis, MN.
- Tharp, J.A., Y-Bham, D., 1980. *A Rhade–English Dictionary with English–Rhade Finderlist*, Australian National University, Canberra, Australia.
- Tincoff, R., Jusczyk, P., 1996. Are word-final sounds perceptually salient for infants? Poster Presented at the Fifth Conference on Laboratory Phonology. Evanston, IL.
- Treiman, T., Danis, C., 1988. Syllabification of intervocalic consonants. *Journal of Memory and Language* 27, 87–104.
- Trubetzkoy, N.S., 1931. Die phonologischen systeme. *Travaux du Cercle Linguistique de Prague* 4, 96–116.
- Trubetzkoy, N.S., 1969. *Principles of Phonology* (C.A.M. Baltaxe, Trans.), University of California Press, Berkeley, CA (original work published 1939).
- Tryon, D.T., 1967a. *Dehu Grammar*, Australian National University, Canberra, Australia.
- Tryon, D.T., 1967b. *Dehu–English Dictionary*, Australian National University, Canberra, Australia.
- Turkina, E., 1982. *Latviesu–Angļu Vardnīca: Ap 30 000 Vardu*, Avots, Riga, Latvia.

- Valdman, A., Klinger, T.A., Marshall, M.M., Roltet, K.J., 1988. Dictionary of Louisiana Creole, Indiana University Press, Bloomington, IN.
- Velten, H., 1943. The growth of phonemic and lexical patterns in infant language. *Language* 19, 281–292.
- Vihman, M.M., 1996. *Phonological Development: The Origins of Language in the Child*, Blackwell, Cambridge, MA.
- Vihman, M.M., Ferguson, C.A., 1987. The acquisition of final consonants. In: Viks, U. (Ed.), *Proceedings of the Eleventh International Congress of Phonetic Sciences*. Tallinn, Estonia, USSR, pp. 381–384.
- Webster, M., 1964. *The New Merriam-Webster Pocket Dictionary*, Simon and Schuster, New York.
- Welleman, B.L., Case, I.M., Mengert, K.G., Bradbury, D.E., 1931. Speech sounds of young children. *University of Iowa Studies in Child Welfare* 5, 1–82.
- Wiering, E., Wiering, M., 1994. *The Doyayo Language*, University of Texas at Arlington, Dallas, TX.
- Wu, C., 1998. *Daur*, Lincom Europa, Munich, Germany.
- Zamuner, T.S., 2003. *Input-based Phonological Acquisition*, Routledge, New York.
- Zamuner, T.S., Gerken, L.A., 1998. Young children's production of coda consonants in different prosodic environments. In: Clark, E.V. (Ed.), *Proceedings of the Twenty-ninth Annual Child Language Research Forum*. Center for the Study of Language and Information, Stanford, CA, pp. 13–25.
- Zamuner, T.S., Ohala, D., 1999. Preliterate children's syllabification of intervocalic consonants, In: Greenhill, A., Litterfield, H., Tano, C. (Eds.), 1999. *Proceedings of the 23rd Boston University Conference on Language Development*, vol. 2. Cascadilla Press, Somerville, MA (pp. 753–763).
- Zipf, G.K., 1935. *Psychobiology of Language: An Introduction to Dynamic Philology*, MIT Press, Cambridge, MA (1965 reprint).