Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference “Dialogue 2026”

June 24–26, 2026

Positional alignment of consonant cluster variability in early speech including effect of morphology

Anastasia Chuprina
Institute for Automated Language Teaching and Assessment (ALTA)
University of Cambridge
a.o.chuprina@gmail.com

Abstract

This study explores morphological and positional variation in the production of consonant clusters in early development. An analysis of child speech transcriptions and their adultlike idealised variants is provided in English and French to investigate cluster diversity. A measure of entropy is used to calculate how stable the production of clusters is across word positions (initial, intervocalic, final) and morphological conditions (morphological vs. non-morphological). Results reveal similar positional asymmetries between the child, or a constrained system, and the adultlike, or articulatorily unconstrained variant. Furthermore, morphological complexity destabilises the phonological system at the global scale, illustrated by higher entropy in the final morphological position in English but not French. Finally, gradual convergence between the two systems seems to be prompted by an increase in syntactic complexity of the child speech. Overall, the results are consistent with the statistical learning account of language acquisition.

Keywords: child speech, system stability, morphology, word position, consonant clusters
DOI: 10.29003/2075-7182-2026-24-58-66

Позиционная вариативность стечений согласных в ранней речи, включая роль морфологии

Анастасия Чуприна
Институт автоматизированного обучения и оценки языковых навыков
Университет г.Кембриджа
a.o.chuprina@gmail.com

Аннотация

Это исследование посвящено морфологическим и позиционным различиям стечений согласных на ранних этапах развития речи. В нём проанализирована транскрипция детской речи и её идеализированный квазивзрослый вариант на материале английского и французского языков. Чтобы оценить стабильность фонологической системы использовалась мера энтропии, рассчитанная по позициям кластера внутри слова (начальная, межгласная, финальная) и морфологическим условиям (лемма и словоформа). Результаты обнаруживают позиционную асимметрию, которая оказалась выстроена почти одинаково для детской и квазивзрослой систем. При этом вариативность в детской речи была ожидаемо выше. Также, морфология увеличивает нестабильность системы на глобальном уровне, что подтверждается повышенной энтропией в конечной позиции морфологического условия в английском языке, но не во французском. Наконец, сближение информационной структуры двух систем может быть связано с усложнением синтаксического компонента речи.

Ключевые слова: детская речь, системная стабильность, морфология, позиция в слове, стечения согласных

1 Background

1.1 Variation in learner system

A complex motor skill emerges slowly by means of adaptation of the system to the constraints of the environment. Initially, learners show greater variation in the process of minimising the probabilities of the undesirable outcome (Bernstein, 1967). Following Nikolai Bernstein's model of motor control, initial greater diversity in performance is bound to take place because motor constraints are temporarily relaxed. This gives additional degrees of freedom to the learner to perform the task while slowly learning to manage a larger number of parameters (e.g. in free climbing task, (Cordier et al., 1994)).

Vocal learners, as humans and songbirds, naturally start vocalising promptly after birth gradually showing adaptation to their native or species-specific vocalisations (Saffran et al., 1996; Doupe and Kuhl, 1999; Lipkind et al., 2020). Crucially, variation is observed in several domains behind vocal production –motor performance, signal acoustics and size of the repertoire. Variation in behaviour gradually, through the initial exploration, starts to serve adaptative and efficient motor activity (Hadders-Algra, 2018). Along with variability in motor control, acoustics of the bird's plastic song or of infant's babbling varies in noise and range to the degree greater than the crystallised song (Marler and Peters, 1982), (Liu Wan-Chun and Fernando, 2009) or the consecutive stage of word production (Oller, 1973; Kuhl and Meltzoff, 1996)). Given the favourable developmental conditions, the repertoire of the vocal elements and their acoustic properties will naturally narrow down to the adultlike expected variation. However, there seems to be another driver of the reduction specific to the communicative signal itself, namely a sequence. Vowels produced in a sequence come to gradually occupy a more limited region of the vocal tract, leading to smaller vowel space, than those produced singly in infants of $3 - 12$ months of age, illustrating the process of exploration (Goldfield, 2000). And in birds, early phase of overproduction is followed by selective attrition of the vocal repertoire, limiting the original vocabulary 4-5 times the original numbers in the plastic song (Peters and Nowicki, 2017).

Thus, we expect an overall higher uncertainty in the early production of consonant clusters, a particular feature of phonological complexity present in many but not all languages of the world, than in its ideal, or articulatorily unconstrained, variant. We will analyse clusters by their position within a word and their morphological condition.

1.2 Positional effects

Positional differences in the phonetic realisations have been observed diachronically, synchronically, in speech processing, and in language acquisition. For example, voicing, as in the transition of *life* from Latin *vita* into Spanish *vida* or French *vie*, has been observed in the intervocalic position (Odden, 2005). In synchrony, (Hutin et al., 2022)'s results show that voicing alternations by position vary even between languages of a single family and edge positions differ from the intervocalic one. Experimentally, the results of artificial language learning, (Finley and Newport, 2010) and word recognition (McQueen, 1998)) show that adults use perceptual cues and knowledge about positional phonotactic constraints of their language for word segmentation. The effect of position has also been observed in bilingual child speech –(Grimm and Domahs, 2025) show that 5 year-old bilingual speakers of German produced word-initial clusters with high accuracy in contrast to word-final clusters, which were more susceptible to the influence from another language.

In addition to the overall variation exemplified above, clusters differ along an additional dimension –morphology. Indeed, a cluster that occurs after inflecting a verb like [kt] in *asked* has different probabilities from [kt] in *tact*, which might lead to different learning strategies for the two levels, lexical and morphological (Dressler, 2006; Schwartz et al., 2024). In the following section some predictions will be outlined according to the positional and morphological condition of cluster production.

1.3 Is position and its morphological condition a constraint for learning?

Based on the above, entropy will be used to account for the stability within the system of speech production in early language development for English and French. Given articulatory limitations of a developing speaker, we expect the phonetic structure to be less stable than in the adultlike system, resulting in higher entropy.

Second, we can expect positional differences but we do not yet have a clear picture of their alignment at the level of the system. Although we know that languages differ in their preferences for the cluster location –e.g. in Italian they are mostly attributed to the word initial position or they can occupy all positions as in Polish but be restricted based on sonority, an articulatory component and morphological complexity (Zydorowicz and Orzechowska, 2017)– we do not know about their distributions.

Finally, an empirical question is whether final position of a word will reflect competing pressures between the morphological constraint, thus reducing the entropy, or an opposite effect in case of morphology offering increased combinatorial opportunities.

2 Materials & Methods

2.1 Data sample

The data were sourced from CHILDES database (version 2021.1, (MacWhinney, 2019)), using R library *childesr* (Braginsky et al., 2019). Only the corpora that contained IPA-formatted phonetic transcriptions of both variants, child's utterances (e.g. *he's green* [his gwin]) and their model or adultlike variant ([hiz ɡɹiːn]) were included. We used data from children with two or more recording sessions (See Table 1). The children were recorded at their homes while they engaged in normal play activities with their caretakers. The corpora contain annotations for word token, its lemma and part of speech. Only nouns and verbs were included into the study.

Language	N of children	Age range	Mean (sd) loquacity in tokens
English	6	12–48	4305 (2836)
French	5	12–52	3219 (1979)

Table 1: Description of the sample.

Two languages of the dataset (French: Lyon corpus (Demuth and Tremblay, 2008) and Yamaguchi corpus (Yamaguchi, 2012); English: Providence corpus (Demuth, 2004) have comparable phonological and morphological profiles in terms of consonant inventories, consonant-to-vowel ratios, a complex syllabic structure, concatenative formatives of grammatical markers (Dryer and Haspelmath, 2013). However, they differ in how phonology interacts with morphological structure. While French has phonological alternations that in rare cases result in a consonant cluster, like *vouloir-voudrais* [vudʁɛ] or *regarder-regarde* [ʁəɡaʁd], English concatenation of grammartical morphemes regularly results in a cluster as in *paint-paints* [peɪnts], *grill-grilled* [ɡɹɪld, *drive-drives* [dɹaɪvz], *do-don't* [doʊnt].

2.2 Preprocessing

Before searching for clusters, a custom-made algorithm in Python programming language cleared the transcriptions off of the diacritics in order to account for the phonemic inventory instead of gradient vocal production. Then it recorded consonantal combinations in word-onset, intervocalic, and word-final positions. The clusters that were combinations with glides (e.g. mj or kw) or most likely denoted co-articulated consonants (e.g. t͡p) were excluded as well as clusters produced in isolation. The full list of IPA characters of cluster constituents can be found in Appendix A.

The dataset was limited to the words of up to four-syllable length, based on the number of the vocalic elements in a word, which comprised $>$ 95% of child productions and $>$ 97% in adultlike system, with 1-syllable words accounting for $>$ 50% of the sample. Medial position was assigned to all intervocalic clusters with no regards to the word token length. The majority of clusters consisted of two consonants ($>$ 90% of all clusters) in both systems, child and adultlike phonologies.

We quantified the degree of formal divergence between surface forms and their lemmas by computing the Levenshtein distance (Levenshtein, 1966), a measure of the minimum number of single-character insertions, deletions, or substitutions required to transform one string into another. The *stringdist* function in R (method = ”lv”) was used. Based on these values, we then categorised clusters into two groups: Non-morphological condition where tokens have a Levenshtein distance of 0 (e.g.*bird-bird*), indicating that the gloss and stem were identical, and Morphological condition that included tokens with a Levenshtein distance greater than 0 (e.g. *bird-birds*).

Table 2 provides the overall token and type numbers of consonant clusters across languages and morphological conditions. The entropy values are known to be sensitive to the size and distribution of these

Language	Morphology	Child N Clusters	Child Unique Clusters	adultlike N Clusters	adultlike Unique Clusters
English	NM	21904	626	33828	175
	M	10211	519	13572	152
French	NM	4437	222	7632	86
	M	1907	156	3870	68

Table 2: Number of cluster tokens and unique cluster types produced by children and adultlike speakers across languages and morphological conditions, NM is non-morphological condition.

parameters. To obtain size-unbiased entropy estimations, the clusters were binned into equally-sized bins to allow the two systems, child and adultlike, to differ on their complexities only and thus to ensure comparability of information content across developmental stages. The bin size varied between 10 clusters per bin to the entirety of cluster production. Binning ensured that each child had series of equally-sized bins of clusters. The size of 50 clusters per bin is chosen to be reported here because of two reasons. First, the results of modelling were the same with the binning of greater and smaller sizes. Second reason is pragmatic as in both child and adultlike datasets, the rate of cluster production per position per child in a given month was observed with median of 16–31 and a mean of 33–64 clusters. Thus, a bin size of 50 captures enough data to compute reliable entropy, which aligns well with the actual production of clusters across positions and children. Binning was performed for each position. The resultant age range per position was nearly identical: Initial $14-58$, Medial $12-55$, Final $14-54$ months of age.

We also calculated proxy measures for vocabulary and syntax. Vocabulary was estimated based on the number of unique lemmas produced in a session and then averaged per bin. Similarly was calculated the mean length of utterance (in tokens).

2.3 Diversity estimation

Entropy was estimated via R package *entropy* (Hausser and Strimmer, 2008). The literature mentions a variety of methods (Chao et al., 2013), (Arora et al., 2022). We used four bias-corrected estimators: Dirichlet–Jeffreys (Krichevsky and Trofimov, 1981), Dirichlet–Schürmann–Grassberger (Schürmann and Grassberger, 1996), Miller–Madow (Miller and Miller, 1955), and Chao–Shen (Chao and Shen, 2003). All measures were highly correlated (Pearson r > .92), indicating that relative differences across systems were robust to estimator choice. We report results based on the Schürmann–Grassberger Dirichlet estimator, which has previously been applied to phoneme and phonological distributions in child speech (Sainburg et al., 2022). It provides alphabet-size–dependent smoothing appropriate for moderately sparse categorical data. By contrast, the Chao–Shen estimator, deemed as the most reliable one (Moscoso Del Prado Martín, 2017), is more commonly used for word-level distributions with a large number of unseen categories. We modelled the entropy using both, the Dirichlet–Schürmann–Grassberger and Chao-Shen estimators. The pattern of results was consistent across both of them.

2.4 Analyses

All analyses were conducted in R (R Core Team, 2024). The response variable, entropy, represents the amount of information in bits in cluster distribution for each position. We fitted linear mixed-effects models (R package *lme4*, (Bates et al., 2015) in two steps: First, we measured cluster diversity as a function of language, morphological condition, word position, syntax and vocabulary, separately in child and in adultlike data; then, we evaluated diversity in the adultlike data, and, finally, modelled whether the difference between child and adultlike diversities ($D = H_{adult} - H_{child}$) responds to the change in vocabulary and/or syntax.

A linear mixed-effects model was fitted with a by-child random intercept to account for repeated observations and a by-child random slope for age to capture individual developmental trajectories. Session-level measures of mean length of utterance (MLU, in tokens), loquacity in a session (in tokens), and

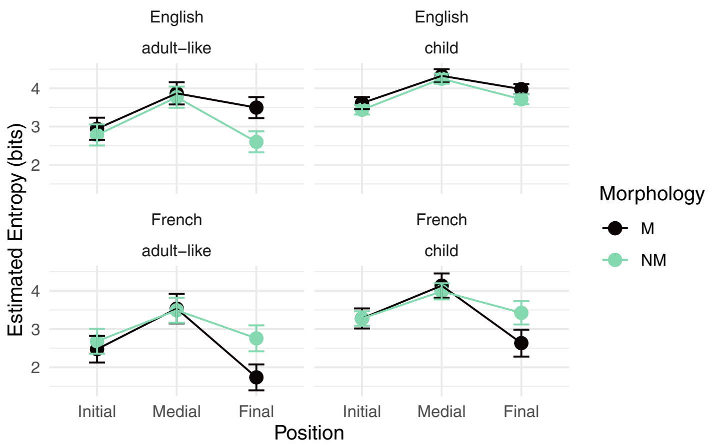

Figure 1: Estimated entropy (in bits) by cluster position across two languages and their morphology, separately for the child and the adultlike production system. Points show the estimated marginal means from linear mixed-effects models, with lines connecting positions and error bars representing 95% confidence intervals.

vocabulary (unique token count in a session) were aggregated by child and age bin using the median, and merged with the entropy dataset to provide aligned covariates for modeling. Together with age, these predictors were standardized (mean = 0, SD = 1) prior to analysis. While entropy was calculated on fixed-size bins, the mean age of each bin was used as the predictor in the model. Tukey post hoc tests via R package *emmeans* (Lenth and Piaskowski, 2017) were used in case of significance of the main effects in the model.

The structure of random effects in the final model was chosen on the basis of likelihood-ratio tests (χ^2) and AIC comparisons (Akaike, 1998). Residual diagnostics was performed to assess linearity and homoscedasticity, and to check for model misspecification. R package *ggplot2* (Wickham, 2016) was used to generate figures.

3 Results

Here we describe morphological patterns in positional variation in child system (3.1), in developmentally unconstrained system (3.2). We then evaluate whether vocabulary gains and/or syntactic complexity affect the convergence between the systems (3.3).

3.1 Stability in child system

Entropy showed neither clear developmental trend as we did not observe a stable effect of age, nor an effect of syntactic complexity. However, vocabulary was positively associated with entropy ($Est. = 0.06, SE = 0.03, p = 0.01$). We did not observe an independent effect of morphology. Position had significant main effect and in interaction with morphology and language. As can be seen in Figure 1, both languages display a greatest variability in intervocalic position. As for the interaction between morphology and position, the pattern of variability was different for French and English and only for the final position. In French, in morphological condition entropy was lower than in the non-morphological ($Est. = -0.99, SE = 0.24, p < 0.001$). In English, on the opposite, entropy was higher in the final morphological position ($Est. = 0.24, SE = 0.07, p < 0.001$).

3.2 Stability in adultlike system

Entropy of adult-system shows a clear developmental trend: with age overall entropy increases ($Est. = 0.12, SE = 0.02, p < 0.001$) and there is a positive association with vocabulary ($Est. = 0.07, SE = 0.03, p = 0.02$). However, MLU did not have an influence.

As in the child system, we did not observe an independent effect of morphology. All the other key

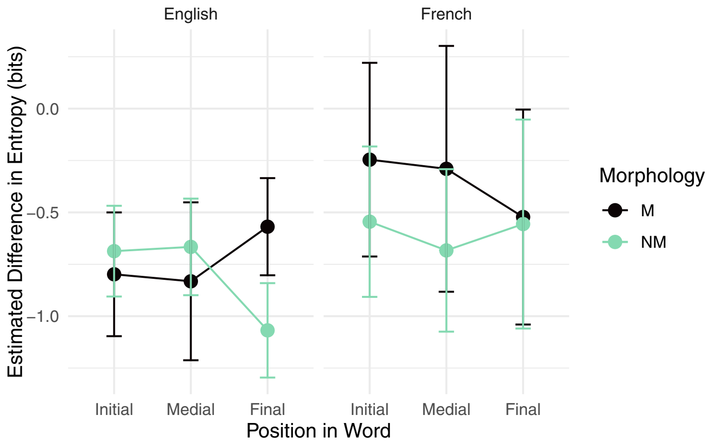

Figure 2: Predicted population-level trajectories of the entropy difference between adultlike and child systems (D). Predictions were generated from the linear mixed-effects model with random effects set to zero (population-level effects).

effects revealed the same pattern as the child data. Position had significant main effect and in interaction with morphology and language: In the final position the effect of morphology was more pronounced. In French, morphological condition for the final position was lower than in the non-morphological ($Est. = -1.5, SE = 0.14, p < 0.001$). In English final morphological position was higher in the non-morphological ($Est. = 0.74, SE = 0.06, p < 0.001$).

3.3 Difference between systems

To examine developmental convergence, we modeled the difference between adult and child entropy values ($D = H_{adult} - H_{child}$) using a mixed-effects model with random intercepts and age slopes by child. A negative intercept ($Est. = -0.55, SE = 0.09, p < 0.001$) indicated that children produced significantly higher entropy values than adultlike system. Neither age nor vocabulary predicted the entropy. However, MLU had a positive effect on difference between the systems, suggesting that convergence towards the stable state was dependent on the syntactic development of a learner ($Est. = 0.17, SE = 0.08, p = 0.03$). Finally, final morphological position of English showed a greater proximity to the adultlike system than the non-morphological condition ($Est. = 0.5, SE = 0.09, p < 0.001$) as illustrated by Figure 2.

4 Discussion

In this study we accounted for the stability of the phonological system in early language development. For this, we used entropy of consonant clusters produced in three positions within a word and tested a possibility of the morphological effect on positional distributions. We observed a greater uncertainty in consonant cluster production among children than in the corresponding idealised system, supporting the relaxation of the articulatory constraint in early development. The findings support the idea that a greater entropy in learner behaviour is an adaptive property of developing systems, proposed by Nikolai Bernstein (Bernstein, 1967) and later work on vocal development. Greater uncertainty of early phonological organisation may reflect a necessary exploratory phase in which learners test and reorganise articulatory, perceptual, and linguistic constraints before stabilising towards adultlike patterns. In future studies we could therefore investigate whether entropy functions as a general developmental marker.

Importantly, the child system structurally mirrored the adultlike system, showing a similar positional asymmetry, albeit less predictable. Both systems revealed morphological effect on the cluster distribution within the final position in the word. In English, morphological change was accompanied by higher instability while in French, the opposite –by greater predictability. For English final position, the observed pattern is in line with our prediction that morphological change may destabilise phonological distribution.

For French, the pattern diverges. The difference might be attributed to the nature of the morphological change. In French, the cluster occurs as a result of a change within the word ending as e.g. [ʁd] in the pair *regarder-il regarde* where the cluster moves between intervocalic and final positions or as e.g. [dʁ] in *attendre-il attend* where the cluster disappears; in English, the cluster is always the result of a collated ending as [ks] in e.g.*look-he looks*. A theoretical question can therefore be raised of how language-specific structure shapes developmental stability. Thus far, the results suggest that morphology may either stabilise or diversify phonological production depending on how morphophonological alternations are organised in a given language. Extending the analysis to diverse languages would help determine whether the observed English–French contrast reflects typological differences across and within language families.

Finally, the findings suggest that linguistic growth reshapes stability, rather than chronological maturation alone. This is in line with the studies of the interconnectedness among the levels of the language system and their possible inseparability in early development (Moscoso del Prado, 2014; **?**). In the future, these results should be validated on the acoustic measures of speech production and through a combination of entropy with articulatory or neural measures of development to better understand how stability emerges across interacting systems during learning.

Data availability

The data presented in this study are available on the Open Science Framework (OSF) at https://osf.io/a6jb2/overview.

References

Hirotogu Akaike, 1998. *Information Theory and an Extension of the Maximum Likelihood Principle*, P 199–213. Springer New York, New York, NY.

Aryaman Arora, Clara Meister, and Ryan Cotterell. 2022. Estimating the entropy of linguistic distributions. // *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, P 175–195, Dublin, Ireland. Association for Computational Linguistics.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1).

Nikolai Bernstein. 1967. *The Coordination and Regulation of Movements.* Pergamon Press.

M Braginsky, A Sanchez, and D Yurovsky. 2019. childesr: Accessing the ‘childes’database. *R package version 0.1*, 2.

Anne Chao and Tsung-Jen Shen. 2003. Nonparametric estimation of shannon’s index of diversity when there are unseen species in sample. *Environmental and Ecological Statistics*, 10(4):429–443.

Anne Chao, Y. T. Wang, and Lou Jost. 2013. Entropy and the species accumulation curve: a novel entropy estimator via discovery rates of new species. *Methods in Ecology and Evolution*, 4(11):1091–1100, November.

Patrick Cordier, Michel Mendés France, Jean Pailhous, and Philippe Bolon. 1994. Entropy as a global variable of the learning process. *Human Movement Science*, 13(6):745–763, December.

Katherine Demuth and Annie Tremblay. 2008. Prosodically-conditioned variability in children’s production of french determiners. *Journal of Child Language*, 35(1):99–127, February.

Katherine Demuth. 2004. Childes english providence corpus.

Allison J. Doupe and Patricia K. Kuhl. 1999. Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, 22(1):567–631, March.

Wolfgang U Dressler. 2006. Introduction: Natural morphology. *Folia Linguistica*, 40(1part2):1–6.

Matthew S. Dryer and Martin Haspelmath. 2013. *WALS Online (v2020.4)*. Zenodo.

Sara Finley and Elissa L. Newport. 2010. Morpheme segmentation from distributional information.

Eugene C Goldfield. 2000. Exploration of vocal tract properties during serial production of vowels by full term and preterm infants. *Infant Behavior and Development*, 23(3–4):421–439, March.

Angela Grimm and Ulrike Domahs. 2025. The acquisition of consonant clusters and word stress by early second language learners of german: Evidence for cross-linguistic influence? *Linguistic Approaches to Bilingualism*, 15(2):212–242, February.

Mijna Hadders-Algra. 2018. Early human motor development: From variation to the ability to vary and adapt. *Neuroscience Biobehavioral Reviews*, 90:411–427.

Jean Hausser and Korbinian Strimmer. 2008. entropy: Estimation of entropy, mutual information and related quantities.

Mathilde Hutin, Martine Adda-Decker, Lori Lamel, and Ioana Vasilescu. 2022. When phonetics meets morphology: Intervocalic voicing within and across words in romance languages. // *Interspeech 2022*, P 3438–3442. ISCA.

R. Krichevsky and V. Trofimov. 1981. The performance of universal encoding. *IEEE Transactions on Information Theory*, 27(2):199–207, March.

Patricia K. Kuhl and Andrew N. Meltzoff. 1996. Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, 100(4):2425–2438, October.

Russell V. Lenth and Julia Piaskowski. 2017. emmeans: Estimated marginal means, aka least-squares means, October.

Vladimir I Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals. // *Soviet physics doklady*, volume 10, P 707–710.

Dina Lipkind, Andreea Geambasu, and Clara C. Levelt. 2020. The development of structured vocalizations in songbirds and humans: A comparative analysis. *Topics in Cognitive Science*, 12(3):894–909.

Wada Kazuhiro Liu Wan-Chun and Nottebohm Fernando. 2009. Variable food begging calls are harbingers of vocal learning. *PLoS ONE*, 4(6):e5929.

Brian MacWhinney. 2019. Understanding spoken language through talkbank. *Behavior Research Methods*, 51(4):1919–1927, August.

Peter Marler and Susan Peters. 1982. Developmental overproduction and selective attrition: New processes in the epigenesis of birdsong. *Developmental Psychobiology*, 15(4):369–378.

James M. McQueen. 1998. Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, 39(1):21–46.

G Miller and G Miller. 1955. Note on the bias of information estimates.

Fermín Moscoso Del Prado Martín. 2017. Vocabulary, grammar, sex, and aging. *Cognitive Science*, 41(4):950–975, May.

Fermin Moscoso del Prado. 2014. Grammatical change begins within the word: Causal modeling of the co-evolution of icelandic morphology and syntax. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 36(36).

David Odden. 2005. *Introducing Phonology*. Cambridge University Press, 1 edition, February.

D. Kimbrough Oller. 1973. The effect of position in utterance on speech segment duration in english. *The Journal of the Acoustical Society of America*, 54(5):1235–1247, November.

Susan Peters and Stephen Nowicki. 2017. Overproduction and attrition: the fates of songs memorized during song learning in songbirds. *Animal Behaviour*, 124:255–261, February.

R Core Team, 2024. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Jenny R. Saffran, Richard N. Aslin, and Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science*, 274(5294):1926–1928, December.

Tim Sainburg, Anna Mai, and Timothy Q. Gentner. 2022. Long-range sequential dependencies precede complex syntactic production in language acquisition. *Proceedings of the Royal Society B: Biological Sciences*, 289(1970):20212657, March.

Geoffrey Schwartz, Katarzyna Dziubalska-Kołaczyk, and Radosław Święciński. 2024. Phonotactic and morphonotactic influences on the (a)synchronicity of consonant clusters in polish. *Studia Linguistica Universitatis Iagellonicae Cracoviensis*, 141(3):205–217, August.

Thomas Schürmann and Peter Grassberger. 1996. The predictability of letters in written english. *Fractals*, 04(01):1–5, March.

Hadley Wickham. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Naomi Yamaguchi. 2012. *Parcours d'acquisition des sons du langage chez deux enfants francophones.* Theses, Université de la Sorbonne nouvelle - Paris III, May.

Paulina Zydorowicz and Paula Orzechowska. 2017. The study of polish phonotactics: Measures of phonotactic preferability. *Studies in Polish Linguistics*, 12(2).

A IPA Characters of Constituents

In Child-Produced Clusters
's', 'f', 'ʝ', 'ɣ', 'ð', 'k', 'ɾ', 'p', 'v', 'n', 'ʧ', 'r', 'ɱ', 'ʔ', 'ʁ', 'ʙ', 'ɕ', 'ɡ', 'z', 'b', 'ŋ', 'ʦ', 'c', 'm', 'l', 'ʒ', 'ɲ', 'ʃ', 'ç', 'h', 'χ', 'ʀ', 'd', 't', 'θ'
In Ideal Speaker Clusters
'ʤ', 's', 'f', 'ð', 'k', 'p', 'v', 'n', 'ʧ', 'r', 'ʔ', 'ʁ', 'ɡ', 'z', 'b', 'ŋ', 'm', 'l', 'ʒ', 'ʃ', 'h', 'ʀ', 'd', 't', 'θ'