languages

MDPI

Review

Speech Segmentation with Prosodic and Statistical Cues Is Language-Specific in Infancy

Mireia Marimon [1,*], Amanda Saksida [2,3], Barbara Höhle [4] and Alan Langus [4,*]

[1] Center for Brain and Cognition, Department of Engineering, University Pompeu Fabra, 08020 Barcelona, Spain
[2] Pediatric Audiology and Otolaryngology Unit, Institute for Maternal and Child Health—IRCCS "Burlo Garofolo", 34137 Trieste, Italy; amanda.saksida@gmail.com
[3] Centre for Discourse Studies, Educational Research Institute Ljubljana, 1000 Ljubljana, Slovenia
[4] Department of Linguistics, University of Potsdam, 14476 Potsdam, Germany; hoehle@uni-potsdam.de
* Correspondence: mireia.marimon@upf.edu (M.M.); alanlangus@gmail.com (A.L.)

Abstract

Speech segmentation is one of the first tasks infants face when learning their mother tongue. It has been argued that statistical learning could function as a gateway to speech segmentation in the absence of pre-existing knowledge about the language to be acquired. However, infants also segment speech with prosodic cues, such as lexical stress. Here, we review recent evidence from studies that look at how infants weigh statistical and prosodic information when segmenting continuous speech. We argue that the idea that statistical regularities have a main role in early speech segmentation, as evidenced in English-learning infants, is not found with German-learning infants. With more natural speech stimuli, German-learning infants only become sensitive to statistical regularities in the speech signal by their first birthday. We provide further support for this hypothesis by showing that there are cross-linguistic differences in how statistical models segment child-directed speech (CDS) and that CDS changes as infants grow. This suggests that speech input to younger infants is not tailored for speech segmentation with statistical cues, but that it is subject to cross-linguistic differences like prosody.

Keywords: transitional probability; statistical learning; child-directed speech; speech segmentation; language-specific

Academic Editors: Mariapaola D'Imperio and Sónia Frota

Received: 7 February 2025
Revised: 27 August 2025
Accepted: 3 September 2025
Published: 19 September 2025

Citation: Marimon, M., Saksida, A., Höhle, B., & Langus, A. (2025). Speech Segmentation with Prosodic and Statistical Cues Is Language-Specific in Infancy. *Languages*, *10*(9), 240. https://doi.org/10.3390/languages10090240

1. Introduction

Speech directed at infants, like speech to adults, unfolds as a sequence of sounds where word boundaries are not marked by pauses. Therefore, a critical part of language acquisition involves parsing this fluent speech into meaningful units that correspond to the words of the language. Infants have been shown to rely on multiple cues to segment speech, including allophonic variation (Jusczyk et al., 1999b), phonotactic patterns (Mattys & Jusczyk, 2001; Mattys et al., 1999), coarticulation cues (Johnson & Jusczyk, 2001), statistical cues between syllables (Aslin et al., 1998) and prosodic cues (Morgan & Saffran, 1995; Jusczyk et al., 1999a; Houston et al., 2000; Höhle et al., 2009; Marimon et al., 2024). A key question in understanding speech segmentation is how infants weigh these different cues. Some studies have investigated whether one cue (e.g., statistical vs. prosodic) dominates the other and whether reliance on one emerges earlier in development. Evidence suggests that the weighting of these cues can change over time, reflecting developmental shifts in how infants prioritize different sources of information (Thiessen & Saffran, 2003). In this

manuscript, we review recent evidence on the role of prosodic and statistical cues in speech segmentation and discuss how cue-weighting evolves during early language development, considering new cross-linguistic data.

1.1. Prosody in Speech Segmentation

There is growing evidence that infants are sensitive to prosodic cues from birth. For example, newborn infants can discriminate languages based on phrasal prosodic cues alone (Nazzi et al., 2000; Ramus et al., 2000), they are sensitive to prosodic boundary cues (Christophe et al., 2001), lexical stress (Sansavini et al., 1997) and can even segment streams of syllables alternating in prominence (Fló et al., 2019). While the former two are related to phrasal prosody, which provides broad rhythmic and intonational patterns, the other two are related to word-level stress patterns, which vary systematically across languages and constitute more language-specific segmentation cues. Throughout this manuscript, we refer to the latter. For example, while some languages, like German, English, and Dutch, have regular lexical stress on the first syllable in disyllabic words, other languages, such as Portuguese, can have variable lexical stress (Cunha & Cintra, 1984), and languages like French lack lexical stress at the word level (Cutler & Mehler, 1993; Féry et al., 2011). Crucially, also the acoustic realization of lexical stress varies across languages: While some languages rely on pitch and intensity to signal prominence in words, others tend to rely more on vowel duration. This raises the question to what extent young infants, who have only limited exposure to the language that surrounds them, use prosodic cues to parse continuous speech into possible word candidates.

Adult listeners show robust language-specific processing of prosodic cues, such as stress (Cutler, 2005; Dupoux et al., 1997). Research shows that language-specific perception of prosody emerges during the first year of life. For example, sensitivity to lexical stress has been attested in Italian newborns (Sansavini et al., 1997), in 6-month-old Spanish-learning infants (Skoruppa et al., 2011), in 6-month-old German-learning infants (Höhle et al., 2009), in French-learning infants from 4 to 10 months of age (Friederici et al., 2007; Höhle et al., 2009; Skoruppa et al., 2011) and in 5-month-old Portuguese-learning infants (Frota et al., 2020). This sensitivity to lexical stress becomes language-specific during the first year of life. For example, while German-learning infants show a preference for the dominant trochaic (strong-weak) stress pattern (Höhle et al., 2009), Portuguese-learning infants show a preference for an iambic (weak-strong) pattern (Frota et al., 2024) and French-learning infants show no preference for either trochaic or iambic stress pattern (Höhle et al., 2009). This suggests that the early sensitivity to the acoustic realization of speech prosody develops into a preference for the native language prosodic structure sometime during the first year of life and that it depends on exposure to the native language.

There is also evidence that both infants and adults can segment words from continuous speech using native language prosody. For example, English-learning 7-month-old infants and adult English listeners can segment disyllabic word forms that match the prosodic pattern typical of their native language (i.e., trochaic stress pattern) (Morgan & Saffran, 1995; Jusczyk et al., 1999b; Houston et al., 2000; Cutler & Norris, 1988). Similar findings have also been observed in infants acquiring other stress-timed languages. For example, German-learning infants segment trochaic words from a continuous artificial speech stream at 6 and 9 months (Marimon et al., 2022; 2024). Dutch-learning infants begin segmenting trochaic words at 10 months, failing to do so at 7.5 months (Kuijpers et al., 1998; Kooijman et al., 2009). In contrast, 8-month-old Canadian-French infants (a predominantly phrase-final stress language) failed to segment English words from Canadian-English passages and Canadian-English infants failed to segment words from Canadian-French passages (Polka & Sundara, 2012). Parisian French-learning infants fail to segment disyllabic words from

natural speech passages until 12 months of age (Nazzi et al., 2006). These studies show that the segmentation strategies that infants use to find words from continuous speech are specific to the native language prosody already during the first year of life.

In short, while research suggests that prosody can help young infants to find possible word candidates in continuous speech during the first year of life, their ability to do so is language-specific in several ways. First, while prosody appears to facilitate speech segmentation in stress-based languages like English, Dutch and German, it can fail to do so in syllable-based languages like French (Marimon et al., 2025). Second, even in stress-based languages, the emergence of speech segmentation with lexical stress shows variability, with segmentation emerging considerably earlier in infants acquiring German and English than Dutch. This raises the question of how young infants, who have limited experience with their native language, discover the dominant prosodic patterns of their native language and how they might start segmenting words from continuous speech.

1.2. Statistical Cues in Speech Segmentation

To solve the segmentation problem without pre-existing knowledge of the specific language to be segmented, research has focused on statistical learning. Corpus studies of English adult speech have demonstrated that the probability of co-occurrence between adjacent syllables tends to be higher when the syllables occur within words than when they occur across word boundaries (Harris, 1955; Hayes & Clark, 1970). Transitional Probabilities (TPs) are typically calculated using the formula: $TP = P(Y \mid X) = Frequency(XY)/Frequency(X)$. For example, in the English phrase "pretty baby", the syllable "pre" is more often followed by the syllable "ty" than the syllable "ty" is followed by the syllable "ba", which could be followed by any syllable from the next word. This means that the TP between "pre" and "ty" is higher than the TP between "ty" and "ba" (Saffran et al., 1996; Aslin et al., 1998). Research in speech segmentation in young infants suggests that infants exploit these differences in TPs between syllables to find possible word candidates in continuous speech.

Both infants and adults can segment words from a continuous stream of syllables based on statistical information, i.e., TPs (e.g., Hayes & Clark, 1970; Saffran et al., 1996; Aslin et al., 1998; for a review, see Saffran & Kirkham, 2017). Infants can use the relative difference of TPs to segment continuous sequences of linguistic (artificially synthesized speech: Saffran et al., 1996; natural speech: Pelucchi et al., 2009) and non-linguistic stimuli (Endress & Mehler, 2009; Kirkham et al., 2002), and their ability to compute TPs is correlated with a variety of linguistic skills later in language development (Arciuli & Simpson, 2012; Kidd & Arciuli, 2016; Marimon et al., 2022; Graf Estes et al., 2007; Obeid et al., 2016; Saffran, 2001; Erickson et al., 2014). The ability appears to be present already at birth (with speech stimuli, Teinonen et al., 2009; Fló et al., 2019; non-speech stimuli, Bulf et al., 2011; Kudo et al., 2011) and it is evolutionarily so primitive that it has been observed in several species of non-human animals (Toro & Trobalón, 2005; Hauser et al., 2001). Therefore, it has been argued that speech could be segmented, at least in part, through domain-general statistical computations that have not specifically evolved for language processing and that do not depend on pre-existing knowledge of the language to be acquired (Saffran et al., 1996).

While TPs have emerged as a key mechanism through which infants may solve the speech segmentation problem, this ability has several important limitations, mainly because the stimuli used in these experiments are rather unnatural and insufficiently complex. For example, infants can readily segment continuous speech when all words are of equal length (e.g., all words are CVCV), but they struggle when word length varies (e.g., half of the words are CVCV and the other half are CVCVCV) (Johnson & Tyler, 2010). Similarly, when infants are primed with disyllabic words, they are more likely to segment disyllabic

words from continuous speech but fail to segment trisyllabic words. Conversely, when primed with trisyllabic words, they will segment those but not disyllabic words (Lew-Williams & Saffran, 2012). Since word length varies widely in natural speech (Saksida et al., 2017), these findings raise questions about how well infants' ability to segment artificial syllable sequences generalizes to more complex natural speech. In fact, evidence suggests that speech segmentation using TPs is most successful with artificially synthesized speech stimuli (Black & Bergmann, 2017), which may facilitate segmentation by promoting rhythmic predictions of word boundaries (Marimon et al., 2022). This raises the question of how well TPs can be applied to segment entirely naturalistic linguistic input.

1.3. Cue Weighting

Studies of speech segmentation in young infants that look at individual speech segmentation cues suggest that infants are equipped with a toolkit of various speech segmentation abilities that include both distributional as well as prosodic cues. There is some evidence that infants can use multiple cues to speech segmentation when more than one cue signals word boundaries. For example, when exposed to a stream of syllables where both TPs and lexical stress signal word boundaries, infants segment speech better when they can rely on both cues (e.g., Johnson & Jusczyk, 2001, Experiment 4). While this suggests that infants track the occurrence of multiple cues at once, there is also some evidence that young language learners do not weigh different cues for segmentation equally (e.g., Johnson & Jusczyk, 2001; Mattys et al., 1999). In fact, experiments where infants are familiarized with speech stimuli where statistical and prosodic cues signal different word boundaries have challenged whether TPs can help young infants to segment speech without pre-existing linguistic knowledge.

For example, Mattys et al. (1999) familiarized 9-month-old English-learning infants with a stream of syllables containing lexical stress and phonotactic cues pitted against each other. The syllable stream consisted of CVC-CVC disyllabic non-words that were stressed either on the first or on the second syllable. Crucially, the adjacent consonants at syllable boundaries had either a low probability of occurring at word boundaries and a high probability of occurring inside of words in English speech or vice versa. The results show that infants listened significantly longer to disyllabic nonce words with a trochaic (strong-weak) pattern that violated phonotactic cohesion compared to iambic (weak-strong) nonce words that adhered to phonotactic cohesion. Similar findings were also obtained by Johnson and Jusczyk (2001), who familiarized 8-month-old English-learning infants with a syllable string that contained conflicting cues to word boundaries, namely lexical stress and TPs. After familiarization, infants were tested with words based on prosodic cues and words based on TPs. They found that English-learning 8-month-olds' segmented the familiarization stream with prosodic rather than statistical cues. These findings suggest that, shortly before their first birthday, English-learning infants rely more on prosodic than on distributional cues to segment continuous speech into possible word candidates.

This preference for prosodic cues over statistical cues when segmenting speech is subject to developmental constraints. Using a similar experimental design as Johnson and Jusczyk (2001), Thiessen and Saffran (2003) showed that, while English-learning 7-month-olds relied more strongly on statistical cues in their segmentation performance, 9-month-olds were more strongly guided by prosodic cues. Based on these findings, the authors argue for an initial dominance of statistical cues over prosodic cues. They explain this change in cue relevance by a crucial difference in the status of the cues. As opposed to the dominant lexical stress pattern in spoken languages, which are language specific, computing TPs in the speech input needs no specific language knowledge and therefore may serve as an initial gateway to speech segmentation. This developmental shift was

subsequently also supported by findings showing that 5-month-old English infants rely more on statistical cues (Thiessen & Erickson, 2013) and that 11-month-olds rely more strongly on prosodic cues (Johnson & Seidl, 2009).

However, the idea that TPs universally serve as a gateway to speech segmentation in the absence of acquired linguistic knowledge was recently questioned by cue-weighting studies with German-learning infants (Marimon et al., 2022, 2024) and French-learning infants (Marimon et al., 2025). They tested 6-month-old German-learning infants following a similar experimental design as previous studies with English-learning infants (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003). Results showed that, when lexical stress and TPs signal different word boundaries, German-learning infants at 6 months rely more strongly on prosodic cues. In addition, when infants were presented only with TPs, they failed to segment the speech stream. Crucially, there was no evidence for a developmental shift from TPs towards prosody in the acquisition of German (Marimon et al., 2022). Instead, at 9 months of age, some German-learning infants relied on prosodic cues and others used TPs to segment the identical linguistic string. These findings do not rule out a role for TPs in early segmentation but suggest that a developmental shift for German-learning infants, if any, would happen in the opposite direction: Infants would start using prosodic cues at early stages in development and then would start to integrate other cues such as TPs in their segmentation strategies. In the case of French, where the language rhythm is syllable-timed and stress is less prominent, statistical cues are expected to play an important role in early segmentation. However, recent findings suggest that even in French, the use of TPs at 6–7 months is modulated by prosodic cues that mark perceptual salience. Specifically, Marimon et al. (2025) found that French-learning infants succeeded at segmenting words based on TPs only when an intensity cue (e.g., increased loudness) signaled different word boundaries than TPs. In contrast, when a duration cue (e.g., syllable lengthening) marked word boundaries that conflicted with those defined by TPs, infants failed to segment the speech stream accordingly. This asymmetry suggests that not all prosodic cues exert the same influence, and that perceptually salient cues like duration may override or interfere with statistical word segmentation. These results highlight the complex interplay between statistical and prosodic cues and underscore that segmentation performance in French-learning infants is not solely driven by TPs, but by the interaction between multiple cues.

Overall, these experiments show that the way young infants weigh different cues when segmenting speech is not only subject to developmental constraints but is also influenced by cross-linguistic differences even in prosodically similar languages like English and German. While there is substantial evidence that infants' speech segmentation abilities become language-specific within the first year, much less is known about how young infants discover the most effective segmentation strategies from exposure to their native language, which likely involves integrating multiple cues. German-learning infants become sensitive to TPs between syllables considerably later than English-learning infants (Marimon et al., 2022, 2024). Although further research is needed, this suggests that the usefulness of TPs for segmenting words from continuous speech may vary across languages and might be less universal than previously assumed, depending on the suitability of these cues within a given language.

Importantly, our aim is not to argue for a strict temporal primacy of prosody over statistical cues, but to highlight that language-specific prosodic preferences, particularly at the word level, can emerge quite early in development—sometimes even at birth. Newborns have been shown to be sensitive not only to prosody (e.g., Sansavini et al., 1997; Moon et al., 2013), but also to statistical regularities (Fló et al., 2019; see also Maye et al., 2002). What our findings suggest is that prosodic cues may be more robust and more salient than TPs depending on the properties of the ambient language. A related conceptual issue

is how "success" in using TPs is defined. Our argument does not assume that infants achieve perfect segmentation; even partial segmentation can support the gradual building of lexical representations. While TPs can provide useful segmentation cues, infants' partial success in using them suggests that their role may be incremental rather than all-or-none. In addition, the segmentation of words using TPs also appears to be influenced by the specific input infants receive (e.g., Stärk et al., 2022; Thiessen & Saffran, 2003). This, again, raises the question whether computation of TPs can indeed provide a language-independent gateway to speech segmentation and it can help infants to discover the lexical stress pattern of their native language. Future research could therefore focus on natural speech perception, cross-linguistic differences between languages, examine how multiple cues signal word boundaries, and explore how infants integrate these cues to solve the segmentation problem.

1.4. Cross-Linguistic Differences in Infant- and Child-Directed Speech

The idea that an effective use of TPs for word segmentation may be dependent on some language-specific adjustment is supported by corpus analyses of child-directed speech (CDS). For instance, CDS in Italian is best segmented using a segmentation algorithm that calculates Forward TPs—that is, the probability that syllable Y will follow syllable X in a given syllable pair XY. In contrast, Hungarian CDS is more effectively segmented using Backward TPs, which calculate the conditional probability that syllable X will precede syllable Y in the syllable pair XY (Gervain & Guevara, 2012). These cross-linguistic differences extend to other languages as well. For example, when comparing CDS across nine languages, Saksida et al. (2017) observed considerable variability in how Forward and Backward TPs segment the speech. The most effective segmentation strategy was correlated with the rhythmic nature of each language. This suggests that speech segmentation using TPs depends on the language being segmented.

Further evidence against the hypothesis that TPs serve as a gateway to speech segmentation comes from a recent corpus analysis in which we examined how the statistical structure of CDS changes as children grow (Langus et al., 2019). The way caregivers speak to their children evolves with age. For example, speech directed at younger children tends to feature more repetition of words and simpler sentence structures (Fernald & Morikawa, 1993; Raneri et al., 2020; Tal et al., 2022), which can influence the distribution of syllables across utterances. High repetition may increase word-internal TPs but reduce opportunities for infants to compute transitional dips at word boundaries, thereby altering the cues available for segmentation. Simpler structures may further concentrate co-occurrence patterns within familiar phrases, potentially shifting the statistical landscape learners are exposed to. Our hypothesis is that, if TPs were a universal gateway to speech segmentation, we would expect speech to younger children to exhibit less cross-linguistic variability and to favor a single segmentation algorithm (e.g., Forward TPs or Backward TPs). To test this hypothesis, we analyzed CDS from six different languages (German, Dutch, Italian, Hungarian, English, and Estonian) transcribed in the CHILDES database (MacWhinney, 2000; available at https://childes.talkbank.org/ (accessed on 20 January 2025)). We focused on how distributional cues used for statistical speech segmentation are influenced by parents' tendency to adjust their speech according to the child's age. For each utterance in each corpus, we determined the age of the child at which it was recorded (from 1 to 60 months). As a result, the original grand corpora were parceled into smaller ones (a total of 196) roughly corresponding to individual CHILDES' sessions of a child and its caregivers (see Table 1 for a detailed composition of the corpora). In contrast to previous corpus analyses that aggregated CDS from children of different ages (cf. Saksida et al., 2017; Gervain & Guevara, 2012), we treated children's age as a continuous factor. Note that

the English corpus is the only one to include transcripts from infants under 4 months of age, which introduces a sampling bias that may limit direct comparisons of early input across languages.

Table 1. Composition of the corpora.

Language	Children/Transcript	Youngest/Oldest (Months)	Sentences (SD)	Word Tokens (SD)	Word Types (SD)	Syllable Tokens (SD)	Syllable Types (SD)
Dutch	2 7	19.20 36.30	291.60 (140.00)	1306.00 (652.00)	293.90 (83.60)	1694.10 (819.00)	318.40 (74.10)
English	29 29	1.50 4.00	374.20 (214.90)	1252.30 (778.90)	229.00 (98.90)	1469.30 (910.40)	258.70 (107.60)
Estonian	3 14	19.80 49.10	178.80 (123.20)	776.80 (604.70)	261.10 (148.10)	1228.40 (887.40)	255.10 (113.40)
German	2 63	10.00 59.10	106.80 (60.00)	538.10 (267.0)	24.80 (73.90)	750.60 (352.20)	250.90 (75.60)
Hungarian	5 57	32.20 59.90	251.60 (189.50)	937.80 (747.20)	344.70 (227.00)	1592.20 (1279.7)	386.00 (215.90)
Italian	3 14	16.10 40.30	186.50 (162.60)	926.50 (754.00)	307.10 (183.70)	1719.30 (1454.0)	233.30 (96.10)

The amount of linguistic input provided to children at various ages was roughly comparable. We found no significant increase in the number of syllable tokens, word tokens, or sentence tokens in CDS as children grew older across the six languages. Despite similar token frequencies, we observed a significant increase in speech complexity. First, words directed at older children contained, on average, more syllables, and sentences directed at older children contained, on average, more words (Figure 1E,F). Additionally, there was a significant increase in syllable and word types with age across all languages (Figure 1A,B). We also observed a significant increase in the type/token ratio for words, but not for syllables (Figure 1C,D). This supports the assumption that words are repeated more frequently to younger children cross-linguistically. However, this repetition pattern was observed only for words, not for syllables. This suggests that, while more syllable types enter the linguistic input, the frequency of repetition for each syllable type remains constant across age groups. We speculate that this reflects a fundamental aspect of the combinatorial nature of spoken language, where a limited set of syllables are combined to form a much larger number of words.

The increase in variability on the word level observed in CDS cross-linguistically was mirrored by a language-specific improvement in the performance of segmentation algorithms. Specifically, we compared Forward and Backward TPs using two types of segmentation algorithms. First, we used a relative algorithm that assigns word boundaries when the TP between two syllables decreases relative to neighboring syllable pairs. Second, we used an absolute algorithm that assigns word boundaries based on the average TPs at word boundaries, using this value as a threshold for segmentation (cf. Saksida et al., 2017). Performance below chance suggests that the segmentation cue is not relevant as the algorithm is yielding more nonwords than actual words. For both the relative and absolute algorithms, when comparing Forward and Backward TPs, we found that segmentation performance with Forward TPs increased significantly with age in English, Estonian, Italian, and Dutch, while performance with Backward TPs improved in Hungarian and German (see Figure 2). However, segmentation performance with the relative algorithm never resulted in more actual words than non-words (see also Saksida et al., 2017). In contrast, the absolute algorithm using Forward TPs outperformed Backward TPs in English, Estonian, Italian, and Dutch at 20 months, while Backward TPs outperformed Forward TPs in German and Hungarian at 30 months. This suggests that, even with relatively little linguistic input, the absolute algorithms can achieve considerable

segmentation performance, although the direction of the TP computation varies depending on the language in question. Note that, while TPs are used here as a way of quantifying statistical structure in the input, we do not take a position on whether infants compute probabilities *per se* or segment via alternative mechanisms such as chunking (e.g., Perruchet, 2019; Jessop et al., 2025). More generally, our focus is on evaluating TPs as a mechanism for early segmentation, rather than adjudicating between TP- and chunking-based models, so our arguments should not be interpreted as a broader rejection of distributional or sequence-based accounts of segmentation (Perruchet, 2019; Frank et al., 2010). Our interpretation focuses on how regularities align with behavioral outcomes, independent of the underlying cognitive process.

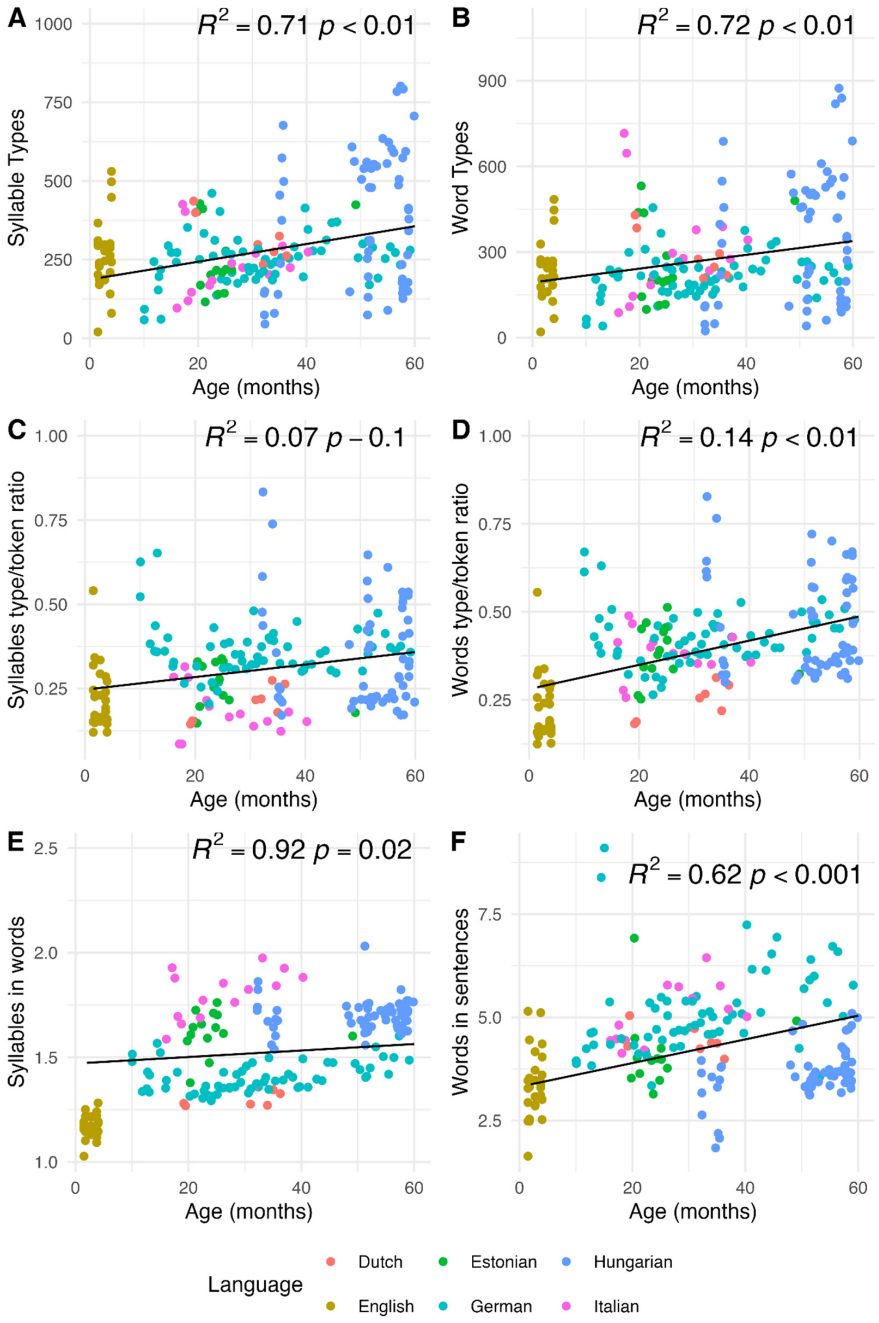

Figure 1. The complexity of child-directed speech in the six languages. (**A**) The number of different syllable types across ages. (**B**) The number of different word types across age. (**C**) Syllable type/token ratio across ages. (**D**) Word type/token ratio across ages. (**E**) The average number of syllables in words across ages. (**F**) The average number of words in sentences across ages.

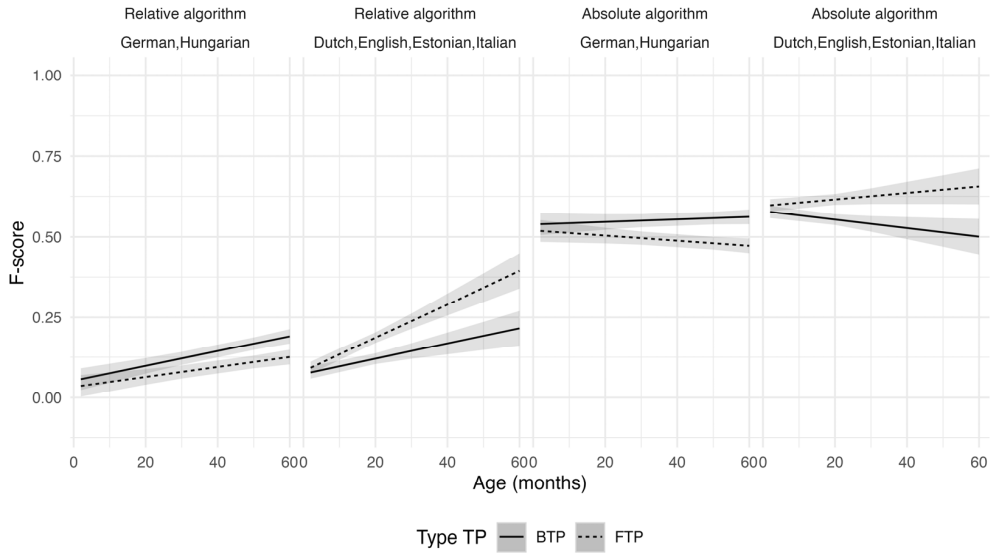

Figure 2. The performance of the Relative and Absolute segmentation algorithms. The change in F-score (Harmonic mean of Precision and Recall) of the Relative and Absolute algorithms with children's age using Backward TPs (BTPs) and Forward TPs (FTPs) in languages where BTPs outperform FTPs (i.e., German, Hungarian) and in languages where FTPs outperform BTPs (i.e., Dutch, English, Estonian, Italian). Shaded areas represent 0.95 confidence intervals. Whenever the proportion of hits is substantially lower than the proportion of falsely selected or missed words, the F-score is lower than 0.5.

By comparing the CDS corpora, we showed that, as children get older, their caregivers' utterances become more complex in several ways: Irrespectively of the language under question, they contain more syllable and word types, have higher word type/token ratios, and have words containing more syllables and sentences containing more words. Note that, even though more different syllable types enter the input as children get older, the different types of syllables are repeated roughly the same amount across ages, possibly because they are combined in different ways in the increasing number of words present in caregivers' input. These results show that, cross-linguistically, caregivers' utterances to their children change gradually as children become older. Utterances to younger children are characterized by sentences that contain fewer words, words that contain fewer syllables, but more repetitions of individual words. While such repetition may enhance local distributional coherence, it does not necessarily improve the overall structure of TPs across the speech stream. As children grow older, caregivers' utterances become statistically more informative, in terms of global co-occurrence patterns between syllables, so that statistical speech segmentation algorithms extract more correctly segmented words from input to older children. This suggests that, despite the potential local benefits of repetition, the simplified speech input to younger children may in fact limit the availability of the broader statistical information needed for effective TP-based segmentation when compared to speech directed to older children. Therefore, the universal tendency to simplify speech to infants seems to impoverish the statistical information in speech input necessary for infants to successfully segment words from continuous speech using statistical cues. Instead, the co-occurrence statistics of syllables appear to show that systematic cross-linguistic differences only become more pronounced in CDS as children become older.

The presented corpus analysis can only demonstrate that TPs in CDS input are not constant but change with input complexity, providing less information to younger children, which may challenge the view on TP primacy. However, the limitations of TPs observed in our analyses should not be taken as evidence that TPs play no role in early segmentation, but rather that their contribution may be relative and incremental. Note also that the pre-

sented transcripts are neither prosodically nor uniformly annotated, and for many of them, no freely available audio recordings exist. As a result, it was not possible to reconstruct the actual prosodic structure of CDS, including the relative strength of cues signaling prominence compared to statistical learning cues. This limitation means that the corpus analysis cannot be informative about prosodic cues primacy. Our conclusions concern segmentation processes in the absence of explicit prosodic information, and they may therefore not fully represent the role that such cues play in naturalistic language acquisition. Relatedly, prior work (e.g., Beech & Swingley, 2023) has shown that phonological variation, including prosodic variation, can significantly affect segmentation outcomes in computational models, underscoring the need for future research incorporating prosodically annotated or audio-accessible corpora.

2. Conclusions

Research suggests that infants possess a toolkit of mechanisms to identify potential word candidates in continuous speech. However, a detailed analysis of speech prosody and statistical regularities between syllables indicates that these cues are language-specific and must be acquired by infants during the first year of life. Our review further highlights that the weighting and integration of these cues differ across languages and developmental stages. These differences underscore the importance of cross-linguistic and developmental research to fully understand how infants learn to segment speech.

Author Contributions: Conceptualization, M.M. and A.L.; methodology, M.M., A.S. and A.L.; formal analysis, A.S. and A.L.; data curation, A.S. and A.L.; writing—original draft preparation, M.M.; writing—review and editing, A.S., B.H. and A.L.; visualization, A.L.; supervision, B.H. and A.L.; project administration, M.M.; funding acquisition, M.M. and B.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by European Union's Horizon Europe, grant number 101108884 and Deutsche Forschungsgemeinschaft grant number 317633480. The APC was partially funded by Open Access Publishing Fund of the University of Potsdam.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created. The data analyzed are freely available following the CHILDES corpora sharing guidelines at https://childes.talkbank.org/ (accessed on 20 January 2025).

Conflicts of Interest: The authors declare no conflict of interest.

References

Arciuli, J., & Simpson, I. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*(2), 286–304. [CrossRef]

Aslin, R., Saffran, J., & Newport, E. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324. [CrossRef]

Beech, C., & Swingley, D. (2023). Consequences of phonological variation for algorithmic word segmentation. *Cognition*, *235*, 105401. [CrossRef]

Black, A., & Bergmann, C. (2017). Quantifying infants' statistical word segmentation: A meta-analysis. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th annual meeting of the cognitive science society* (pp. 124–129). Cognitive Science Society.

Bulf, H., Johnson, S., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*, 127–132. [CrossRef] [PubMed]

Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, *2*(3), 385–394. [CrossRef]

Cunha, C., & Cintra, L. (1984). *Novo gramática do portugués scontemporaneo*. Ediçoes Joao Sá da Costa.
Cutler, A. (2005). *Native listening: Language experience and the recognition of spoken words*. MIT Press.
Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics*, *21*(1), 103–108. [CrossRef]
Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 113–121. [CrossRef]
Dupoux, E., Pallier, C., Sebastián, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory Language*, *36*, 406–421. [CrossRef]
Endress, A., & Mehler, J. (2009). Primitive computations in speech processing. *The Quarterly Journal of Experimental Psychology*, *62*(11), 2187–2209. [CrossRef]
Erickson, L., Thiessen, E., & Graf Estes, K. (2014). Statistically coherent labels facilitate categorization in 8-month-olds. *Journal of Memory and Language*, *72*, 49–58. [CrossRef]
Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, *64*(3), 637–656. [CrossRef] [PubMed]
Féry, C., Hörnig, R., & Pahaut, S. (2011). Correlates of phrasing in French and German from an experiment with semi-spontaneous speech. In C. Gabriel, & C. Lleó (Eds.), *Intonational phrasing in Romance and Germanic: Cross-linguistic and bilingual studies* (pp. 11–41). John Benjamins.
Fló, A., Brusini, P., Macagno, F., Nespor, M., Mehler, J., & Ferry, A. L. (2019). Newborns are sensitive to multiple cues for word segmentation in continuous speech. *Developmental Science*, *22*, e12802. [CrossRef]
Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*(2), 107–125. [CrossRef]
Friederici, A., Friedrich, M., & Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology*, *17*, 1208–1211. [CrossRef]
Frota, S., Butler, J., Uysal, E., Severino, C., & Vigário, M. (2020). European Portuguese-learning infants look longer at iambic stress: New data on language specificity in early stress perception. *Frontiers in Psychology*, *11*, 1890. [CrossRef]
Frota, S., Severino, C., & Vigário, M. (2024). Unfolding prosody guides the development of word segmentation. *Languages*, *9*(9), 305. [CrossRef]
Gervain, J., & Guevara, E. (2012). The statistical signature of morphosyntax: A study of Hungarian and Italian infant-directed speech. *Cognition*, *125*(2), 263–287. [CrossRef] [PubMed]
Graf Estes, K., Evans, J., & Else-Quest, N. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech Language and Hearing Research*, *50*(1), 177–195. [CrossRef] [PubMed]
Harris, Z. S. (1955). From phoneme to morpheme. *Language*, *31*, 190–222. [CrossRef]
Hauser, M., Newport, E., & Aslin, R. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton top tamarins. *Cognition*, *78*, B53–B64. [CrossRef] [PubMed]
Hayes, J., & Clark, H. (1970). Experiments on the segmentation of an artificial speech analog. In J. Hayes (Ed.), *Cognition and the development of language* (pp. 221–234). Wiley.
Houston, D., Jusczyk, P., Kuljpers, C., Coolen, R., & Cutler, A. (2000). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review*, *7*(3), 504–509. [CrossRef]
Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, *32*(3), 262–274. [CrossRef] [PubMed]
Jessop, A., Pine, J., & Gobet, F. (2025). Chunk-based incremental processing and learning: An integrated theory of word discovery, implicit statistical learning, and speed of lexical processing. *Psychological Review*. Advance online publication. [CrossRef]
Johnson, E., & Jusczyk, P. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*(4), 548–567. [CrossRef]
Johnson, E., & Seidl, A. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, *12*(1), 131–141. [CrossRef]
Johnson, E., & Tyler, M. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339–345. [CrossRef] [PubMed]
Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999a). Infant's sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, *61*(8), 1465–1476. [CrossRef]
Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999b). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*, 159–207. [CrossRef] [PubMed]
Kidd, E., & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development*, *87*(1), 184–193. [CrossRef] [PubMed]
Kirkham, N., Slemmer, J., & Johnson, S. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), 35–42. [CrossRef]

Kooijman, V., Hagoort, P., & Cutler, A. (2009). Prosodic structure in early word segmentation: ERP evidence from Dutch ten-month-olds. *Infancy*, *14*(6), 591–612. [CrossRef]

Kudo, N., Nonaka, Y., Mizuno, N., Mizuno, K., & Okanoya, K. (2011). On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Developmental Science*, *14*(5), 1100–1106. [CrossRef]

Kuijpers, C., Coolen, R., Houston, D., & Cutler, A. (1998). Using the head-turning technique to explore cross-linguistic performance differences. In C. Rovee-Collier, L. Lipsitt, & H. Hayne (Eds.), *Advances in infancy research* (pp. 205–220). Ablex.

Langus, A., Marimon, M., Saksida, A., Boll-Avetisyan, N., & Höhle, B. (2019). *Cross-linguistic evidence for age-related changes in the statistical structure of child-directed speech*. [Manuscript submitted for publication]. Department of Linguistics, University of Potsdam.

Lew-Williams, C., & Saffran, J. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition*, *122*, 241–246. [CrossRef] [PubMed]

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.

Marimon, M., Berdasco-Muñoz, E., Höhle, B., & Nazzi, T. (2025). Use of statistical and acoustic cues for speech segmentation in French-learning 7-month-old infants and French-speaking adults. In *Discoveries in Cognitive Science* (Volume 9, pp. 189–209). Open Mind. [CrossRef]

Marimon, M., Höhle, B., & Langus, A. (2022). Pupillary entrainment reveals individual differences in cue weighting in 9-month-old German-learning infants. *Cognition*, *224*, 105054. [CrossRef] [PubMed]

Marimon, M., Langus, A., & Höhle, B. (2024). Prosody outweighs statistics in 6-month-old German-learning infants' speech segmentation. *Infancy*, *29*(5), 750–770. [CrossRef] [PubMed]

Mattys, S. L., & Jusczyk, P. W. (2001). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 644–655. [CrossRef]

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*, 465–494. [CrossRef]

Maye, J., Weiss, D. J., & Aslin, R. N. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111. [CrossRef]

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica*, *102*(2), 156–160.

Morgan, J., & Saffran, J. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, *66*, 911–936. [CrossRef]

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, *54*, 283–299. [CrossRef]

Nazzi, T., Jusczyk, P., & Johnson, E. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, *43*, 1–19. [CrossRef]

Obeid, R., Brooks, P., Powers, K., Gillespie-Lynch, K., & Lum, J. (2016). Statistical learning in specific language impairment and autism spectrum disorder: A meta-analysis. *Frontiers in Psychology*, *7*, 1245. [CrossRef] [PubMed]

Pelucchi, B., Hay, J., & Saffran, J. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–685. [CrossRef]

Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Topics in Cognitive Science*, *11*(3), 520–535. [CrossRef]

Polka, L., & Sundara, M. (2012). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: Native language, cross-dialect, and cross-language comparisons. *Infancy*, *17*(2), 198–232. [CrossRef]

Ramus, F., Hauser, M., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, *288*, 349–351. [CrossRef] [PubMed]

Raneri, D., Von Holzen, K., Newman, R., & Bernstein Ratner, N. (2020). Change in maternal speech rate to preverbal infants over the first two years of life. *Journal of Child Language*, *47*(6), 1263–1275. [CrossRef] [PubMed]

Saffran, J. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, *81*, 149–169. [CrossRef] [PubMed]

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-olds. *Science*, *274*, 1926–1928. [CrossRef]

Saffran, J., & Kirkham, N. (2017). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203. [CrossRef]

Saksida, A., Langus, A., & Nespor, M. (2017). Co-occurrence statistics as a language-dependent cue for speech segmentation. *Developmental Science*, *20*(3), 1–11. [CrossRef]

Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, *33*(1), 3–11. [CrossRef] [PubMed]

Skoruppa, K., Cristia, A., Peperkamp, S., & Seidl, A. (2011). English-learning infants' perception of word stress patterns. *Journal of the Acoustical Society of America*, *130*, 50–55. [CrossRef]

Stärk, K., Kidd, E., & Frost, R. (2022). Word Segmentation Cues in German Child-Directed Speech: A Corpus Analysis. *Language and Speech*, *65*(1), 3–27. [CrossRef]

Tal, S., Smith, K., Culbertson, J., Grossman, E., & Arnon, I. (2022). The impact of information structure on the emergence of differential object marking: An experimental study. *Cognitive Science*, *46*(3), 1–31. [CrossRef]

Teinonen, T., Fellmann, R., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *Neuroscience*, *10*(1), 21. [CrossRef]

Thiessen, E., & Erickson, L. (2013). Discovering words in fluent speech: The contribution of two kinds of statistical information. *Frontiers in Psychology*, *3*, 590. [CrossRef] [PubMed]

Thiessen, E., & Saffran, J. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716. [CrossRef]

Toro, J. M., & Trobalón, J. (2005). Statistical computations over a speech stream in a rodent. *Perception and Psychophysics*, *67*(5), 867–875. [CrossRef] [PubMed]