Is it easier to segment words from infant- than adult-directed speech? Modeling evidence from an ecological French corpus

Georgia Loukatou (georgialoukatou@gmail.com)
Laboratoire de sciences cognitives et psycholinguistique, Département d'études cognitives, ENS, EHESS, CNRS, PSL University
Paris, France

Marie-Thérèse Le Normand (marielenormand@mac.com)
INSERM & LPP (Laboratoire Psychopathologie et Processus de Santé), Université Paris Descartes, Sorbonne
Paris, France

Alejandrina Cristia (alecristia@gmail.com)
Laboratoire de sciences cognitives et psycholinguistique, Département d'études cognitives, ENS, EHESS, CNRS, PSL University
Paris, France

Abstract

Infants learn language by exposure to streams of speech produced by their caregivers. Early on, they manage to segment word forms out of this continuous input, which is either directly addressed to them, or directed to other adults, thus overheard. It has been suggested that infant-directed speech is simplified and could facilitate language learning. This study aimed to investigate whether features such as utterance length, segmentation entropy and lexical diversity could account for an advantage in segmentability of infant-directed speech. A large set of word segmentation algorithms was used on an ecologically valid corpus, consisting of 18 sets of recordings gathered from French-learning infants aged 3-48 months. A series of textual analyses confirmed several simplicity features of infant-, compared to adult-directed speech. A small segmentation advantage was also documented, which could not be attributed to any of those corpus features. Some particularities of the data invite further research on more corpora.

Keywords: language acquisition; infant-directed speech; computational modeling; word segmentation; unsupervised learning

Introduction

Infants acquire language early on, building a vocabulary of several hundred word forms by 11 months of life (Ngon et al., 2013). Since most word forms do not appear in isolation (Brent & Siskind, 2001), much previous work studies how infants segment (i.e., pull out) forms from their caregivers' running input. A close look at this input shows that it is not homogeneous, but instead contains some speech addressed to the infants themselves (infant-directed speech or IDS) and some speech overheard by infants which is addressed to others, including adults (adult-directed speech or ADS). These two speech registers differ along many dimensions, including some that may impact word segmentation.

Broadly, IDS has been claimed to present properties that would facilitate language acquisition, with IDS being phonologically, syntactically, and semantically simplified (Soderstrom, 2007). Other characteristics are more relevant to word segmentation. First, IDS may have a higher proportion of single-word phrases (Brent & Siskind, 2001), and phrases might be shorter in length (Newport, Gleitman, & Gleitman, 1977) than in ADS. In shorter phrases, more words would occur at phrase edges, which should improve segmentation: Phrase edges, easily perceptible, are word boundaries provided "for free". Indeed, infants may be more successful at recognizing and segmenting phrase-final words (E. Johnson, Seidl, & Tyler, 2014). Additionally, shorter phrases entail that the set of possible segmentations for each phrase is smaller, lowering segmentation ambiguity. For instance, Fourtassi, Börschinger, Johnson, and Dupoux (2013) showed that ADS might be more ambiguous to segment, when comparing an ADS to an IDS corpus. Second, words may be shorter (Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011), which should mean that word, morphemes, and syllable boundaries coincide more often and there are fewer places to posit or miss positing a boundary. Third, there may be more repetitions, therefore fewer hapaxes (words uttered only once), and overall less lexical diversity (Soderstrom, 2007). Low lexical diversity means fewer target words need to be found. There might be more cues to help segment out frequently repeated words, than words that appear rarely or once. Indeed, one computational modeling study found that artificially reducing phrase length and increasing word repetition in a corpus improved word segmentation with one word segmentation model (Batchelder, 1997). Based on these hypotheses and previous work, we predict that the task of recovering wordforms is easier in IDS than ADS.

Naturally, IDS features may not be the same across infant ages. IDS addressed to very young infants may differ from that addressed to older infants, possibly resembling ADS more as infants get older. For example, IDS features may become less accentuated as the infant grows up; repetitions might decrease, utterance length and lexical diversity increase with age (Henning, Striano, & Lieven, 2005; Soderstrom, 2007). According to the hypotheses explained above, IDS addressed to younger infants should be "easier" to segment than IDS to older infants.

In this paper, we aim to address the question of whether it is easier to segment wordforms from IDS than ADS, using multiple word segmentation models, and taking into account changes with infants' age. In the next section, we review previous modeling work more thoroughly, before introducing our own approach.

Previous studies

Some studies tested whether infants learn more from IDS than ADS in an experimental situation. However, improvements for IDS compared to ADS could be due to the fact that infants pay more attention when they listen to IDS, and thus learn more from it. This method cannot reveal whether, above and beyond this attentional effect, there are intrinsic *informational* differences that affect segmentability. Fortunately, there is a complementary method to approach this question with a colder eye, which builds on computational models of word segmentation. The input to such word segmentation models is usually speech transcriptions, in order to control for differences such as attention capture and acoustic implementation. Segmentation models used for this method are based on findings by experimental studies that infants might make use of statistical cues. Computational models of infant word segmentation can be grouped into two conceptual classes: lexical and sublexical. Sublexical models segment based on local cues, such as transitional probabilities and phonotactics. Lexical models build a lexicon based on recurrent chunks of speech identified with Bayesian probabilities or by memorizing isolated words.

Little previous modeling work has specifically compared IDS and ADS. Four representative studies are summarized in Table 1. For these four studies, improved segmentation performance was found for IDS than ADS: 15% for Batchelder (2002), 5-8% for Fourtassi et al. (2013), 2-10% for Ludusan, Mazuka, Bernard, Cristia, and Dupoux (2017) and 3-10% for Daland and Pierrehumbert (2011). A recent paper critiqued this previous work as follows (Cristia, Dupoux, Ratner, & Soderstrom, 2018). IDS mainly involved caregivers addressing their infants during predefined tasks (e.g., a play session in the laboratory) or in short visits to the child's home. In the former case, by constraining the context, the structure and lexicon of caregivers might have been limited and adapted to that task. And in both cases, being observed could affect caregivers' behavior, who might produce less spontaneous and more formal speech. Moreover, ADS was mostly addressed to an unfamiliar person (experimenter). These conversations are likely more formal than ADS between caregivers in daily life, and could increase the complexity of the speech. As shown by E. Johnson, Lahey, Ernestus, and Cutler (2013), IDS differs more from ADS to unfamiliar adults, than ADS to familiar adults. This could result in increased qualitative differences between registers and probably overestimated differences in segmentability.

Indeed, Cristia et al. (2018) recently documented a considerably smaller IDS advantage when modeling segmentation on an ecological English IDS and ADS corpus. The corpus consisted of transcriptions from excerpts of day-long recordings; thus infants' linguistic environment was recorded while they were going on with their daily lives, resulting in realistic IDS and ADS. Across a wide range of lexical and sublexical models, the IDS advantage ranged from -2% to 8%, with only 3 models providing evidence of an advantage greater than a measure of error. Interestingly, the difference between registers was further reduced when IDS was matched to ADS in corpus length.

The present study

We contribute to this literature in three main ways. First, we specifically describe IDS-ADS differences using various corpus description tools. We compare the registers in: phrase length, word length, ratio of single word phrases, intrinsic segmentation ambiguity (using segmentation entropy), lexical diversity (using Moving Average Type-Token Ratio – MATTR–, so as to control for corpus size), and ratio of hapaxes. Some, but not all of these features have been separately looked at in previous studies (i.e. Fourtassi et al., 2013 measured segmentation ambiguity and Batchelder, 1997 measured word and phrase length, repetitiveness). This is the first study to systematically investigate a plurality of language features on the same IDS-ADS corpus. We test whether IDS is simpler than ADS, as far as these features are concerned. Moreover, following Batchelder (2002), we further investigate whether variation in these features can actually account for the segmentability of a register.

Second, IDS corpora coming from a wide infant age range have been used by previous research, but IDS addressed to infants of different ages were, most of the times, merged together. One exception is Batchelder (1997), who documented that IDS to younger children (13-18 months) produced more successful results than IDS to older children (22-25 months), whereas ADS results from mothers of younger versus older infants didn't differ. In this paper, we specifically ask whether some IDS features interact with infant age and whether segmentability of IDS might actually be affected by age. For that, we include IDS and ADS from a wide age range, and further investigate possible correlations between features, segmentation scores, and infant age.

Third, we follow Cristia et al. (2018) by analyzing a completely ecological child-centered corpus, based on excerpts of day-long recordings, and which thus contains natural ADS and IDS as the child hears over the course of the day. The results of our study would provide more evidence to the question whether differences in home-recorded IDS and ADS are smaller than those between less controlled IDS-ADS contrasts (see Table 1).

In addition to these three main contributions, we extend the range of languages studied to European French.

Methods

We segmented IDS and ADS of each infant separately. Scripts used for corpus preprocessing, phonologization, and segmentation as well as results and supplementary material are available at `https://osf.io/6vwse/?view_only=0bc4f6c0e23040cbbb92e26d414d4a7a`. Statistical analyses were carried out in R (R Core Team, 2013).

Table 1: Summary of design in previous modeling studies comparing IDS and ADS segmentation. In Language(s), Eng stands for English, Jap for Japanese, Span for Spanish. Under IDS and ADS, we describe the corpora. The specific corpora used were: R= RIKEN; H= Hamasaki; C= Spontaneous Japanese; BR= Bernstein Ratner; B= Buckeye; D= Deuchar & Clark 1992, Marrero; M= Miyata 1995; novel= Moon and the Sixpense; short stories were written by Alejandro Dolina (MacWhinney, 1996). Under model, we note the type of model used: lex for lexical and sublex for sublexical.

Study	Language(s)	Infant age(s)	IDS	ADS	model
Batchelder (2002)	Eng.	1;1-1;9	play session (BR)	novel	1 lex
Batchelder (2002)	Span.	1;8-8;0	CHILDES (D)	short story	1 lex
Batchelder (2002)	Jap.	1;3-3;1	home play session (M)	science book	1 lex
Daland et al. (2011)	Eng.	various	all CHILDES	interview (B)	1 sublex
Fourtassi et al. (2013)	Eng.	1;1-1;9	play session (BR)	interview (B)	1 lex
Fourtassi et al. (2013)	Jap.	2;2-3;7	play session (H)	lecture (C)	1 lex
Ludusan et al. (2017)	Jap.	1;6-2;0	play session (R)	lecture (C)	1 lex, 3 sublex

Corpus

Sixteen typically developing native French-speaking infants (eight girls, eight boys; ages 3-48 months, M=20, SD=13), whose families were highly educated, were included. Two of the infants were recorded at two different ages. Each child was recorded 10-16 hours per day, three days a week, in their natural environments. The original recordings are available online (Canault, Le Normand, Foudil, Loundon, & Thai-Van, 2016a, 2016b; VanDam et al., 2016). Next, 18 10-min samples, totaling 3 hours per child (1 hour per day), were selected for orthographic transcription by two native French speakers, as detailed in Canault et al. (2016b). The main criteria for selection reported was that a number of activities were sampled, and that there be a high number of productions by the child and the adult. For the present project, the transcriptions of the first day for all infants were corrected by a native French speaker, who made sure that the definition of utterance was stable (and corrected any other errors, such as misattributions or orthographic errors). The coder annotated whether an adult caregiver's utterance was directed to the target child, an adult, or other, using content and context. Utterances addressed to the target child constituted the IDS corpus and those directed to an adult were the ADS corpus.

Pre-processing

Pre-processing was carried out using custom scripts written mainly in bash and in python, available from `https://github.com/georgialoukatou/French_ADS_IDS_segmentation_Lyon`. All extraneous codes (such as punctuation marks or "xxx", the code indicating that what was said could not be understood by the transcriber) were removed, leaving only the orthographic representation of the adults' speech. The corpora were phonologized with the French voice of the espeak TTS system (Duddington, 2012), using the phonemizer wrapper (Bernard, 2018), which further syllabifies according to the Maximum Onset Principle.

Before segmentation, all spaces between words were removed, leaving the input parsed into minimal units. The minimal units were either phones or syllables. Both phonemes and syllables were tested with all models. Utterance boundaries were preserved as such, since they are supposedly salient to infants (Shukla, White, & Aslin, 2011). This constitutes the input to the model. After preprocessing, the 18 infant-directed corpora contained M=487 (SD 350) utterances (range 84 to 1,172 utterances). The 18 adult-directed corpora contained M=238 (SD 230) utterances (range 15 to 780 utterances).

For comparability with previous work, we evaluate the models' performance using lexical token F-scores, measured by comparing the original version of the input (with spaces between words) against the one returned by the model (with spaces in the hypothesized breaks).

Segmentation

Both corpus description and segmentation were carried out using the WordSeg package (Bernard et al., 2018), available from `https://github.com/bootphon/wordseg/`. Due to space limits, the algorithms are only briefly described here. Full technical details can be found in `https://wordseg.readthedocs.io/`. All algorithms are unsupervised, and inspired in infant experimental work.

We used two representatives of the sublexical word segmentation class contains, called DIBS and TP for short. The Diphone Based Segmentation algorithm (DiBS; Daland & Pierrehumbert, 2011) is based on the idea than a phoneme sequence often spanning phrase boundaries would probably span word breaks.

The Transitional Probabilities algorithm family (TP; Saksida, Langus, & Nespor, 2017) is based on the concept that syllable pairs with lower statistical coherence tend to span word breaks. Forward TP (FTP) measures the frequency of occurrence of the syllabic sequence AB given the frequency of occurrence of the syllable A. Backward TP (BTP) measures the frequency of occurrence of the syllabic sequence AB given the frequency of occurrence of the syllable B. The Relative versions (FTPr or BTPr) threshold TPs against that of neighboring sequences. The Absolute versions

Table 2: Paired t-tests measuring feature differences across IDS and ADS. Word length is measured in phonemes. % 1-w phrase stands for ratio of single word phrases. % hapaxes stands for percent of hapaxes. IDS gives the mean values of each feature on the IDS corpus, with standard deviation in parentheses. ADS shows the mean values of each feature on the ADS corpus with standard deviation in parentheses. The window size for MATTR is 10 words. "p" gives the p-value of the t-test.

Feature	IDS	ADS	p
Word length	2.86 (.08)	2.80 (.11)	.071
Phrase length	5.89 (.85)	6.73 (.86)	*
% 1-w phrase	.18 (.06)	.13 (.05)	**
Entropy	.02 (.004)	.03 (.01)	.31
MATTR	.89 (.03)	.93 (.02)	***
% hapaxes	.39 (.22)	.48 (.27)	***

(FTPa or BTPa) instead threshold on the average of all TPs over the sum of different syllable bigrams.

We used two representatives of the lexical class as well: AG and PUDDLE. Adaptor Grammar (AG) uses the Pitman-Yor process, a stochastic process of probability distribution which prefers the reuse of frequently occurring rules versus creating new ones to build a lexicon, then uses that lexicon to parse the input (M. Johnson, Griffiths, & Goldwater, 2007).

Phonotactics from Utterances Determine Distributional Lexical Elements (PUDDLE, Monaghan & Christiansen, 2010) treats each utterance as a lexical item, unless an already stored item is part of this utterance, and the remainders are phonotactically legal. If so, it breaks up the utterance into segments, and the segments would enter the lexicon as new lexical items.

Finally, two baselines were included: Syll=Word treats each syllable as a word and Utt=Word treats each utterance as a word.

Results

We first investigated whether IDS is simpler than ADS in terms of six corpus features that could affect word segmentation, as described in the reasoning above. The results of paired t-tests comparing the registers for each feature are in Table 2, which shows that four out of six features fit our predictions.

We also noticed that IDS size corpus (M=487, SD=350 per child) was significantly larger than the ADS one (M=238, SD=230), based on a t-test with $t(17)=2.63$, $p=0.02$. This may mean that these infants were exposed to more IDS than ADS, similar to what Cristia et al. (2018) found for English.

The performance of all segmentation algorithms for both registers is captured in Figure 1. IDS is easier to segment than ADS when points are above the dotted diagonal line. There was a small IDS advantage for most algorithms, although some showed the opposite effect (DiBSs,

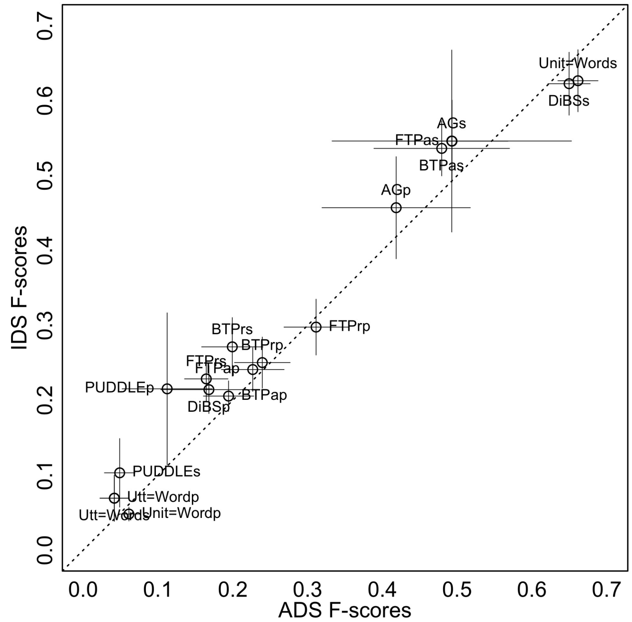

Figure 1: Token F-scores obtained by each algorithm for IDS as function of that for ADS. The final "s" in the model's name means that the basic unit of the corpus was syllables (PUDDLWs, Utt=Words, Unit=Words, DiBSs, FTPas, FTPrs, BTPas, BTPrs, AGs). The final "p" in the model's name means that the basic unit of the corpus was phones (PUDDLWp, Utt=Wordp, Unit=Wordp, DiBSp, FTPap, FTPrp, BTPap, BTPrp, AGp). Error bars show two standard deviations over the 18 corpora.

Unit=Words, Unit=Wordp, FTPrp). We also observe that in many cases the pseudo-confidence intervals cross the diagonal line, suggesting that performance difference is within the range of error. Thus, only FTPrs, BTPrs, Utt=Wordp, PUDDLEp and PUDDLEs showed a clear advantage of IDS. We then tested for overall effects in a linear mixed effect regression model (Bates, Mächler, Bolker, & Walker, 2015) predicting token F-scores from register (IDS or ADS) as a fixed effect, where subject and algorithm (AGs, AGp, DiBSs, DiBSp...) were random effect variables. Register significantly affected token F-scores ($\chi^2(1)=50.87$, $p<.05$, Type II Anova), IDS having a performance advantage of $.03 \pm .004$ (standard error).

Next, we tested whether this performance advantage was due to one of the above-mentioned corpus properties. To see whether performance differences were due to the artifactual difference in corpus length, we also included the number of utterances as a register feature. Thus, 7 new models, each including one of the features as an additional fixed effect, were fit. We then measured the significance of register and features in the new models with a Type II Anova test (Fox & Weisberg, 2011).

If the advantage of IDS was entirely due to one feature, then register would no longer be significant in these addi-

Table 3: Corpus features predict segmentation scores, but do not replace register. β feat stands for the estimated coefficient of that feature; β rgstr for that of register in the new model (which should be compared to 0.03 at the simple model). p features shows whether feature was significant in new model. p rgst shows whether register remained significant in the new model. N. utts stands for number of utterances.

Feature	Feature		Register	
Feature	β	p	β	p
Word length	.02	.48	.03	***
Phrase length	.01	***	.04	***
% 1-w phrase	.06	.29	.03	***
Entropy	-1.58	***	.03	***
MATTR	.5	***	.05	***
% hapaxes	.03	.18	.03	***
N. utts	.00005	***	.02	***

Table 4: Correlation tests (Spearman) of corpus features and infant age for each register. "coef." stands for correlation coefficient.% 1-w phrase stands for ratio of single word phrases. % hapaxes is the ratio of hapaxes.

Feature	IDS coef.	ADS coef.
Word length	.50*	.06
Phrase length	.34	-.56*
% 1-w phrase	-.37	.12
Entropy	-.50*	.70**
TTR	.44	-.37
% hapaxes	.01	.30

tional analyses. Results (in Table 3) showed that phrase length, segmentation entropy, MATTR, and corpus size accounted for variance in the results, but no single feature rendered register effects non-significant.

Next, we investigated whether IDS features change with infant age, with IDS becoming more ADS like as infants age. Spearman correlation tests between properties and infant age for each register separately (Table 4) did not confirm our predictions: Only word length and entropy (neither of which had emerged as register properties on Table 2) correlated with age in IDS; entropy and phrase length did so for ADS. We have no plausible explanation for these effects.

Two infants were recorded twice at different ages, one at 31 and 38 months, the other at 32 and 40 months. Following a recommendation from a reviewer, we inspected these two infants as case studies. An inspection of IDS features demonstrated that phrase length and % of 1-w phrases were the only features having small changes with age, but only the latter would change in the same direction for both infants, increasing by 6% and 1% from the first to the second recording. A few ADS features also changed slightly with age, such as % of 1-w phrases, word length and entropy, but only phrase length changed in the same direction for both infants, decreasing by 1.18 and 1.66 phonemes.

Finally, we created a new model predicting token F-scores register (IDS or ADS) and infant age in months as fixed effects (and model and participant as random effects, as before), and their interaction. Both main effects and the interaction were significant (Age $\chi^2(1)$=4.31, p<.05; Register $\chi^2(1)$=53.14, p<.5; Age:register $\chi^2(1)$=28.81, p<.05). A follow-up analysis separating the registers indicated that ADS scores decreased by .002 ± .0005 (standard error) with age, whereas there was no significant change with age for IDS.

Discussion

In this modeling study, we assessed whether there are informational differences affecting word segmentation between IDS and ADS drawn from the same ecological corpus. First, we investigated whether this naturalistic corpus had IDS-ADS differences in textual features that would make segmentation easier in the former than the latter. We found most features fit our predictions: Phrases were longer, there were more single-word phrases, lexical diversity was lower, and there were fewer hapaxes in IDS than ADS. No significant effect was found for word length and ambiguity. This result contributes to the growing literature documenting IDS features, with the important advantage that current work draws from fully ecological IDS and ADS.

Next, we investigated the segmentability of the corpora using a large set of both lexical and sublexical segmentation models. Although scores varied a great deal across algorithms and some algorithms showed the opposite effect, IDS was overall slightly easier to segment than ADS. The mean difference across registers (CDS minus ADS, in each algorithm separately) was 3%, ranging from –4% to 10%. This effect is smaller than that found in most previous studies, but similar to the one reported by Cristia et al. (2018), who were also drawing from a naturalistic IDS-ADS corpus. This is evidence that previously documented IDS-ADS segmentability differences (as in Table 1) are not representative of what infants actually hear. It is important to note that corpus length across registers was not matched in the present study for practical reasons, but, based on findings by Cristia et al. (2018), we suspect that controling for corpus size would have reduced the IDS advantage even further.

Next, we asked whether some of the above-mentioned textual features uniquely explained segmentability differences across registers. Phrase length, segmentation entropy, and repetitiveness explained significant variance in segmentation scores, above and beyond the effects of register. However, none of the features uniquely explained away the effect of the register, which remained significant in all cases. This means that register effects on segmentability cannot be reduced to any one of these features. Since we only had 18 children's data, we could not fit a model with all 6 features at once for fear of overfitting, but future work with higher power may be able to assess whether these features jointly explain away register, or whether there are other textual features that we have not yet considered.

Furthermore, Canault et al. (2016b)'s corpus allowed us to address a question that has been seldom asked, namely IDS-ADS differences across infant ages. Results of correlations between textual features and age, and a regression model on token F-scores did not support our prediction that IDS would become more like ADS as children aged, and thus the IDS-ADS segmentability gap would close. On the contrary, we found that ADS scores dropped with child age. Although further work is needed, we believe this mainly reflects the lower availability of ADS in children's environment as they age. Indeed, replicating a pattern that had been documented in North American English children (Bergelson et al., 2019), we found the number of ADS utterances dropped for older, compared to younger, children.

Before closing, we would like to acknowledge some limitations of this work. Corpus size was overall small (which may lead to inconsistencies in results; Bernard et al., 2018) and, due to the work involved in collecting daylong recordings and annotating fully spontaneous speech, infant sample size was 18 infants. Moreover, data scarcity was correlated with registers and ages: While only 3 of the 18 IDS corpora contained fewer than 100 utterances, 7 did for ADS, and 4 of those belonged to infants older than 31 months. A decrease of ADS quantities with infant age in such day-long recordings has been documented in previous work on North American English (Bergelson et al., 2019), so it may not be an artifact of the current sample selection. Nonetheless, this trend may entail that if we want to control corpus size, we should over-sample ADS at later ages. However, that may not be necessary for our data, where corpus size failed to explain away the register effect, even though it accounted for some variance beyond registers.

Last, speech transcriptions were used for this study, in an attempt to look for intrinsic informational differences across registers. However, some of the most salient features of IDS are speech-related, such as prosody or intonation and acoustic properties, which might also predict ease of segmentation. Although there is a small literature looking at word segmentation from speech, including comparing IDS and ADS (Ludusan, Seidl, Dupoux, & Cristia, 2015), this task remains extremely challenging for computational modelers, with only one open source model (instantiating a single segmentation strategy) exists, which further limits the value of such a line of research.

In sum, we identified several simplicity features more prevalent in IDS than ADS drawn from an ecological French corpus. We further found a small but significant IDS segmentation advantage, contributing to a recurrent question on the learnability properties of IDS. We showed that the IDS segmentation advantage could not be explained away by any one of those simplicity features, and its size changed with infant age in unexpected directions.

Acknowledgments

References

Batchelder, E. (1997). *Computational evidence for the use of frequency information in discovery of the infant's first lexicon*. Unpublished doctoral dissertation, City University of New York.

Batchelder, E. (2002). Bootstrapping the lexicon: A computational model of infant speech segmentation. *Cognition*, *83*(2), 167–206.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do north american babies hear? a large-scale cross-corpus analysis. *Developmental science*, *22*(1), e12724.

Bernard, M. (2018). Phonemizer [Computer software manual]. `https://github.com/bootphon/phonemizer`, doi = "http://doi.org/10.5281/zenodo.2537809".

Bernard, M., Thiolliere, R., Saksida, A., Loukatou, G., Larsen, E., Johnson, M., ... Cristia, A. (2018). Wordseg: Standardizing unsupervised word form segmentation from text. *Behavior Research Methods*.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33–B44.

Canault, M., Le Normand, M.-T., Foudil, S., Loundon, N., & Thai-Van, H. (2016a). *Lyon homebank corpus.* doi: 21415/T58P6Q

Canault, M., Le Normand, M.-T., Foudil, S., Loundon, N., & Thai-Van, H. (2016b). Reliability of the language environment analysis system (lena$^{\text{TM}}$) in european french. *Behavior research methods*, *48*(3), 1109–1124.

Cristia, A., Dupoux, E., Ratner, N. B., & Soderstrom, M. (2018). Segmentability differences between child-directed and adult-directed speech: A systematic test with an ecologically valid corpus. *Open Mind*, 1–10.

Daland, R., & Pierrehumbert, J. B. (2011). Learning diphone-based segmentation. *Cognitive science*, *35*(1), 119–155.

Duddington, J. (2012). espeak text to speech [Computer software manual].

Fourtassi, A., Börschinger, B., Johnson, M., & Dupoux, E. (2013). Why is english so easy to segment? In *Proceedings of the fourth annual workshop on cognitive modeling and computational linguistics (cmcl)* (pp. 1–10).

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (Second ed.). Thousand Oaks CA: Sage.

Henning, A., Striano, T., & Lieven, E. V. (2005). Maternal speech to infants at 1 and 3 months of age. *Infant behavior and development*, *28*(4), 519–536.

Johnson, E., Lahey, M., Ernestus, M., & Cutler, A. (2013). A multimodal corpus of speech to infant and adult listeners. *The Journal of the Acoustical Society of America*, *134*(6), EL534–EL540.

Johnson, E., Seidl, A., & Tyler, M. D. (2014). The edge factor in early word segmentation: utterance-level prosody enables word form extraction by 6-month-olds. *PloS one*, *9*(1), e83546.

Johnson, M., Griffiths, T. L., & Goldwater, S. (2007). Adaptor grammars: A framework for specifying compositional nonparametric bayesian models. In *Advances in neural information processing systems* (pp. 641–648).

Ludusan, B., Mazuka, R., Bernard, M., Cristia, A., & Dupoux, E. (2017). The role of prosody and speech register in word segmentation: A computational modelling perspective. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 2: Short papers)* (Vol. 2, pp. 178–183).

Ludusan, B., Seidl, A., Dupoux, E., & Cristia, A. (2015). Motif discovery in infant-and adult-directed speech. In *Proceedings of the sixth workshop on cognitive aspects of computational language learning* (pp. 93–102).

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, *7*(3), 185–201.

MacWhinney, B. (1996). The childes system. *American Journal of Speech-Language Pathology*, *5*(1), 5–14.

Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *Journal of child language*, *37*(3), 545–564.

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, id rather do it myself: Some effects and non-effects of maternal speech style.

Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., & Peperkamp, S. (2013). (Non) words, (non) words, (non) words: Evidence for a protolexicon during the first year of life. *Developmental Science*, *16*(1), 24–34.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Saksida, A., Langus, A., & Nespor, M. (2017). Co-occurrence statistics as a language-dependent cue for speech segmentation. *Developmental science*, *20*(3), e12390.

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*, *108*(15), 6038–6043.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532.

VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., & MacWhinney, B. (2016). Homebank: An online repository of daylong child-centered audio recordings. In *Seminars in speech and language* (Vol. 37, pp. 128–142).