The Role of Dialect Variability on Mispronunciation Sensitivity: An Insight to Infants' Early Language Development from a Norwegian Context

Nora R. Serres

Submitted as master's thesis at the Department of Psychology

Faculty of Social Sciences

University of Oslo

Spring 2022

The Role of Dialect Variability on Mispronunciation Sensitivity:
An Insight to Infants' Early Language Development from a Norwegian Context

By: Nora Serres
Advised by: Dr. Julien Mayor & Dr. Natalia Kartushina

Decades of research have highlighted both the differences and commonalities between monolingual and multilingual language acquisition. Yet, substantial differences in the learning environments of monolingual children can modulate learning trajectories towards mature language use. The present study investigates one of the major sources of such difference within monolingual language learning: dialectal variability. Durrant et al. (2015) have shown that dialectal differences have an impact on language acquisition; 20-month-old English infants receiving bidialectal input failed to detect word mispronunciations, a skill mastered by their monodialectal peers. In our study, we investigated mispronunciation sensitivity at a younger age – around 13 months – in both mono- and bidialectal infants. The study was conducted with Norwegian infants, as multi-dialectal input is very common in Norway. The results of the current study will contribute significantly to our understanding of how lexical categories in infants exposed to multiple native dialects develop and inform future work examining the consequences of dialectal exposure on language learning, including how we measure its developmental trajectories.

In this study, 13-month-old participants (n=46) from both monodialectal (n=26) and bidialectal (n=20) households came into the BabyLing lab at the University of Oslo and were tested with an EyeLink eye-tracker. Results from Sequential Bayes Factor analyses reveal that bidialectal infants exhibited word comprehension as well as mispronunciation sensitivity, whereas monodialectal infants did not display a significant effect of naming and can thus not make conclusions about mispronunciation sensitivity (cf Methods). The study has been accepted as a Stage 1 Registered Report in the international journal *Infancy*, with myself as the first author, and work with my advisors has been collaborative. I have led the writing process and cooperated with my advisors to help design the experiment itself based on similar studies conducted in the lab. Plans for the data analysis were a collective effort. Data recruitment, collection, and analysis were also performed by me under supervision of my advisors. This master's thesis is being submitted based on data from the first 46 participants; data collection is still ongoing to meet publication requirements (cf Methods).

Keywords: mispronunciation sensitivity; lexical development; language; dialects; eye tracking

Acknowledgements

First and foremost, I would like to thank my advisors, professors Dr. Julien Mayor and Dr. Natalia Kartushina, for their excellent support and direction. Thank you both for your time, patience, expertise, and willingness to help whenever it was needed. Thank you as well for guiding me through the manuscript process and for helping to problem-solve throughout the experiment and script set-up – I could not have done it without you! I would also like to thank Audun Rosslund for his assistance with the eye-tracker and Ada Koleini for her help in the lab. Thanks as well to Hilde Dragsland who assisted in audio stimuli recording and for friends who allowed me to test the experiment on them before data collection. Further, I would like to thank my parents who have always and continue to support my dreams and interests.

Finally, thank you to all parents and their infants who came and participated in the study!

Table of Contents

Introduction

Typically, language researchers categorize participants based on language exposure and/or proficiency, which often results into contrasting monolingual vs. bi-/multilingual speakers. Researchers have described how processing and learning two or more languages present a challenge to bi- and multilingual children as they need to acquire the phonologies, lexicons and grammars of each language (De Houwer, 2017). Yet, dialectal variability within one language has often been overlooked in research. Although children learning a single language are all monolinguals, there may still be large variability in their exposure to dialectal or regional differences in speech (Bosch & Sebastián-Gallés, 2003; Mattock et al., 2010; Werker & Byers-Heinlein, 2008).

Building lexical representations is at the core of language acquisition, and this process involves a number of phases (Swingley, 2003). One of the first steps is to develop the ability to detect small lexically meaningful changes in the sounds of words, which is critical to being able to build a lexicon containing distinct words that may differ by a single feature such as “sun” versus “fun”. Speech sound representations are constantly being adjusted and refined as the infant is exposed to more linguistic input over time and by different speakers. Thus, an essential component of language learning is to discern which cues mark meaningful differences in their language from ones that mark differences in individual speaker voices, the affect of the speaker, or other linguistically irrelevant differences. This is not an easy task for young language learners: studies have shown that children from 6 months of age are able to recognize words from both familiar and unknown speakers (Bergelson & Swingley, 2018), yet it is not until 10.5 months that infants can recognize familiar words presented in a different affect from the one it was learned in (Singh et al., 2004), suggesting that by this age infants start developing phonological constancy, a capacity to recognize a word despite phonetic variation not altering the phonological structure (Best et al., 2009). Hence, phonological constancy contributes to the establishment and tuning of early word representations.

Accent adaptation

In addition to the above challenges in language development, infants need to adapt to another source of variation stemming from variability in accents. Exposure to multiple accents occurs regularly: people from different regions (e.g., speaking regional dialects) and countries (e.g., foreign-accented speech) meet often, and the need to understand accented speech is crucial for communication. Accented speech typically refers to the pronunciation of words, either native or foreign, that can alter prosodic contour and/or phonetic production of individual segments while preserving established grammatical rules and phonological standards. Variability in lexically irrelevant cues (for instance, in pitch, as typically featured in accented speech) has been suggested to promote infants’ word learning by focusing learners' attention on lexically meaningful cues, as, for example, voice onset time used to distinguish /b/ and /p/ (cf. Rost & McMurray, 2009, 2010).

However, adapting to accents in speech has been shown to take both exposure and maturity. This ability has been demonstrated at early stages of language learning: after brief exposure to a novel accent in a lab setting, 15-month-old Canadian English speakers recognized Australian-accented words in familiar stories (van Heugten & Johnson, 2014), suggesting that exposure to accented speech can facilitate accent adaptation for familiar contexts (lexically-guided adaptation). Van Heugten and Johnson (2017) also conducted another study, where groups of 12.5-month, 14.5-month, and 18-month-old infants, either exposed predominantly to Canadian English or Canadian English plus other/foreign accented speech, were presented with familiar and nonsense words testing preferential listening (of familiar over nonsense words) using a Head-Turn Preference Procedure (HPP). This procedure has the participant seated between two audio speakers, and then a flashing light by one of the speakers gets the participants’ attention before a stream of sound is played from that side until the infant looks away. The streams from each side are either (in this case) known or nonsense words, and the duration of looking towards each side is used to measure their interest in the two stimuli. Their study found that infants exposed predominantly to Canadian English speech in their home environment listened longer to known than nonsense words at 12.5 months already, indicating familiar (phonological) word-form recognition. However, infants with highly variable language exposure revealed word-form recognition at 18 months only, suggesting that exposure from birth to variable input may delay recognition of familiar (phonological) word-forms.

Even so, the results of preferential listening paradigms – showing that infants with exposure to accented speech do not display a preference for familiar words over nonsense/unfamiliar words - can also be attributed to infants’ listening preference. In such paradigms, infants are not offered the possibility of pairing the stimuli to a referent (e.g., an object or its image) and infants exposed to accented speech may have an interest in more variable speech (nonsense words) that competes with their familiarity-based preference for familiar words, translating into reduced looking times towards familiar words. Therefore, it remains unclear whether exposure from birth to accent variability impacts young infants' early word representations.

Dialectal variability

Moreover, many infants nowadays are exposed to dialectal variability, which encompasses accented speech and can include additional sources of variability, characterized by regional variations, featuring intonation and segmental differences (hence, variability in accents), eventual grammatical changes, and even some whole word differences (Venås & Sjekkeland, 2021)[1]. When a child grows up in an environment with two parents speaking different dialects, she must determine which variations are "acceptable" in speech sounds while creating meaningful boundaries between them as they build their phonemic and lexical representations (Bosch & Ramon-Casas, 2011). Norwegian dialects, for example, present all these types of variation: for instance, the word for "not" is either "ikke", "ikkje", "inte", or "itte" depending on the dialect; the /l/ sound can be pronounced with a retroflex flap or without, depending on the dialect as well. Additionally, some specific changes are made such as /v/ to /k/ in words such as "hva" to "ka" and "hvorfor" to "korfor", the Norwegian words for "what" and "why". Thus, dialects provide a unique state of variability, where both lexically relevant (e.g., segmental differences) and irrelevant (e.g., tone) cues can vary between dialects.

The question arises then, how does early exposure to different dialects impact word representations? Specifically, how do children learn to form lexical representations when words are being produced differently across speakers with different dialects, and what implications does this have for early language development in children born and raised in these richly diverse linguistic environments? So far, only a handful of studies have focused on the development of a language's native dialects in monodialectal versus bidialectal infants (Durrant et al., 2015; Floccia et al., 2012; Kartushina & Mayor, accepted pending data collection; Kartushina, Rosslund, & Mayor, 2021; van der Feest & Johnson, 2016; van Heugten & Johnson, 2017).

A few studies on bidialectal infants' language development reveal that early language acquisition in bidialectal infants differs from their monodialectal peers. For instance, Floccia and

---

[1] In the literature, the words "accent" and "dialect" are often used interchangeably. For the sake of consistency and in line with the Norwegian terminology, we use the term "dialects" in the current work.

colleagues (2012) used an Intermodal Preferential Looking paradigm (IPL; Golinkoff et al., 2013), where infants saw two pictures, presented side-by-side on the screen, and were prompted by an auditory stimulus to look at one of them (the target) to examine 20-month-old bidialectal[2] British English toddlers' word representations. In the IPL method, gaze time is measured by an eye-tracker, and difference in time spent looking at the target picture before and after the auditory prompt (naming) is used to measure word comprehension. The authors presented toddlers of parents that either spoke only the local dialect (rhotic) or another (non-rhotic) dialect by naming familiar objects in rhotic or non-rhotic form. In total, picture pairs were presented to each participant, where half of the targets in the test condition were named with a rhotic accent, and the half with a non-rhotic accent. The children in the study were only able to identify stimuli when pronounced in the local (rhotic) form regardless of their parents' dialects, indicating the importance of the local community's speech. Van der Feest and Johnson (2016) also examined specific dialects of a language in slightly older (24 months) infants using IPL, but with a focus on speaker consistency. Dutch infants exposed to either a single dialect that devoices all fricatives or also one that voices some fricatives (i.e., bidialectal) at home received speech input from speakers of each dialect. Van der Feest and Johnson (2016) reported that the infants who received bidialectal input accepted changes in fricative voicing from the different speakers, suggesting flexibility and adaptation to the speakers.

Mispronunciation sensitivity

Durrant and colleagues (2015) used IPL to examine mispronunciation sensitivity - the sensitivity to detect small yet possibly meaningful changes in a word form - in 20-month-old toddlers exposed to monodialectal or bidialectal pronunciations of words at home. Mispronunciation sensitivity is a manifestation of phonological constancy, a phonetic "reference book" that develops in young language learners, which allows them to recognize words in various contexts while discriminating which differences alter the meaning of the word. In Durrant and colleagues' study, toddlers were born and lived in the South West of England; participants whose parents both spoke in this dialect were classified as monodialectal, and participants with parents who spoke different dialects (South West and another) were classified

[2] In the paper the authors use the term "accent" while referring to rhotic vs non-rhotic pronunciation of the "r" sound, but for the sake of consistency we are using the term dialect (as also used in another paper by the same authors (Durrant et al., 2015)), as parents in this study were also representing different dialectal groups of British English.

as bidialectal. The toddlers were presented with picture pairs and targets were either named with correct pronunciations or with a one-feature mispronunciation at either the onset consonant or medial vowel. In their study, Durrant and colleagues found that only monodialectal 20-month-old toddlers showed a difference in recognition between correctly and mispronounced words, as their proportion of looking time at the target image compared to the distractor was higher when the word was pronounced correctly. Bidialectal 20-month-old toddlers, however, looked proportionately longer at the target whether the label was pronounced correctly or not, suggesting that inconsistency in segmental production between parents' dialects impacted phonetic specificity of infants' early word representations.

Note that in monodialectal infants, mispronunciation sensitivity emerges at 7.5 months of age (Jusczyk and Aslin, 1995), and becomes more robust by 11 months of age (Bergelson and Swingley, 2018) and is present for both vowels and consonants by 12 months of age (Mani and Plunkett, 2010; von Holzen and Bergmann, 2021). Yet Durrant et al. 's (2015) study presented above suggests that dialectal variation impacts the degree of specificity of lexical representations in the first two years, also therefore affecting mispronunciation sensitivity. The lack of studies comparing the specificity of lexical representations in mono- and bidialectal infants before 20 months of age does, however, leave a gap in the understanding of lexical development in bidialectal infants and the role of variability in early language learning.[3]

Rost and McMurray's work on word learning (2009, 2010) which examined the role of variability in infants' ability to learn minimal word pairs featuring one phonemic contrast (e.g., /buk/ - /puk/) suggests that variability in non-lexical cues, here in the form of multiple speakers (e.g., through pitch), promotes word learning, whereas variability in lexical cues (e.g., VOT) does not. These results align with research in accents (e.g., rhotic /star/ versus the non-rhotic /sta:/ and formants in vowel production), suggesting that variability in lexical cues can impact word recognition (cf. Durrant et al., 2015). However, when variability features both lexical and non-lexical cues, how do word representations develop? In older bidialectal infants, varied exposure has been shown to benefit learning in multi-dialectal environments. Norwegian 2.5-

[3] Although van Heugten & Johnson (2014) have evaluated the role of accent (not specifically dialect) variability in early language development in young, 12.5-month-old infants, their study used a preferential listening design that in itself cannot reveal the specificity of lexical representations. A lack of preference may simply indicate that infants exposed to accents prefer listening to unfamiliar as much as to familiar words, rather than indicating that they do not recognize familiar words.

year-olds raised in bidialectal home environments learnt words from speech featuring dialects better than their peers raised in monodialectal households, correctly identifying novel pseudowords in a tablet-based e-book more often (Kartushina, Rosslund, and Mayor, 2021). This suggests that long-term (from birth) experience with dialects might benefit later word learning from multi-dialectal input, and further motivates the need to understand the process of lexical representation development before this age. Examining the possible differences in mispronunciation sensitivity is a way to tap into how infant lexical acquisition is influenced by differences in dialectal exposure.

Current Study

The present study explored how infants deal with native dialectal variation, using a mispronunciation paradigm in an IPL task and examining the looking times of both mono- and bidialectal age-matched participants in response to feature changes of presented target words. Knowledge of the specificity of lexical representations that Norwegian infants have at 13 months of age will lead to a better understanding of how both monodialectal and bidialectal infants build their lexical representations and how specific these are around their first year of life: the establishment and tuning of early word representations are beginning to fully form around this age, with the median vocabulary size in comprehension reaching 100 words at 12 months and over 200 words by 14 months of age, already (Mayor & Plunkett, 2011).

In contrast to other European countries, in Norway numerous dialects are still prevalent and widely used today despite the immense changes that languages across Europe have undergone in the last century (Cerruti & Tsiplakou, 2020; Venås & Skjekkeland, 2021). The Norwegian language is generally divided into four broad groups of regional dialects, Eastern (østnorsk), Western (vestlandsk), Central (trøndersk), and Northern (nordnorsk), which differ from each other in their phonological, grammatical, and lexical characteristics (Mæhlum & Røyneland, 2012). Similarly to American versus British English, some specific words (such as the “trunk” versus “boot” of a car) as well as their pronunciation (such as “dog” in British /dɒg/ versus American /dɔ:g/ English) differ while still being mutually intelligible and part of the same language. In addition, in Norway, the Western dialect only has two genders for nouns (masculine -en and neuter -et) while others have three (masculine -en, neuter -et, and feminine -a) and Western and Northern dialects have a characteristic intonational dip at the end of words, whereas the Eastern dialect rises at the end. While most dialects in countries around the world hold stigma attached to them (Dragojevic et al., 2021), regional dialects in Norway are used by people in all social strata and across domains.

The reality of encountering several dialects and variability in speech is thus something that all Norwegian speakers must learn and adapt to from very early on in life. Many children in Norway, and in particular those growing up in the capital city of Oslo, are being born to parents speaking different dialects and thus born into an environment with different linguistic input from each parent. Additionally, parental leave in Norway lasts for an average of 12 months (Norwegian Labour and Welfare Administration, 2021). Consequently, most children are at home during their first year of life, effectively maintaining a significant exposure to their parents' dialects throughout their first year of life and providing us with an opportunity to test infants in mono- and bidialectal households with limited exposure to other environments (e.g., daycare).

In the current study, we examined mispronunciation sensitivity in two groups of Norwegian 13-month-old infants: monodialectal infants receiving similar input from both parents speaking the local Oslo (Eastern) dialect, and bidialectal infants exposed to the Oslo dialect and to a different type of Norwegian dialect (that can belong to one of the remaining three group-types of dialects: Western, Central or Northern). As described above, all four types of dialects are mutually intelligible but clearly distinguishable for their differences in phonetic detail, use of lexical pitch accents, gender attribution for words, and even dialect-specific words (Johnsen, 2012; Mæhlum & Røyneland, 2012). Furthermore, different dialectal input provides variability in both lexically relevant and lexically irrelevant cues, which may impact language learning differently from when infants are exposed to a single dialect.

Mispronunciation sensitivity was tested using an EyeLink eye-tracker to measure looking times (as in Durrant et al., 2015), where infants were presented with pictures and matching word labels, all pronounced similarly across all Norwegian dialects, that were either correctly pronounced or whose onset consonant was mispronounced with a 2-feature change (e.g., *kopp* vs *bopp*). In contrast with Durrant et al. (2015), however, we are examining consonant mispronunciation only; in their study, mispronunciation sensitivity was only detected for vowel mispronunciations, not consonants, possibly attributed to mostly vowel changes across dialects. Our study allows for further assessment of this hypothesis in relation to differences in input, as Norwegian dialects present consonant changes as well (referenced in the Introduction). A larger proportion of baseline-corrected target looking[4] in the post-naming period, in the correct condition compared with the mispronunciation condition, are interpreted as evidence of mispronunciation detection driven by the precision of infant's lexical representations.

Hypotheses

Although to the best of our current knowledge there is no study examining mispronunciation sensitivity in Norwegian 13-month-old infants, recent studies by Durrant and colleagues (2015) and van Heugten and Johnson (2017) suggested that exposure to phonetically variable input (as in infants growing up in bidialectal households) may result in less well-defined lexical representations. The following hypotheses, respecting the requirements of a Sequential Bayes Factor approach (cf Methods) were considered:

Correctly pronounced trials

H1a. We predicted, based on previous research in other languages (Bergelson 2020; Garrison et al., 2020), as well as parental reports (Frank et al., 2016; Simonsen et al., 2014) and an eye-tracking study (Kartushina & Mayor, 2022) in Norwegian, that Norwegian monodialectal 13-month-olds would show word comprehension on correctly pronounced trials, exhibiting a greater proportion of target looks (above 0%) in the post-naming period.

H1b. We predicted that Norwegian bidialectal 13-month-olds would also show word comprehension on correctly pronounced trials, exhibiting a greater proportion of target looks (above 0%) in the post-naming period.

Mispronounced trials

H2a. We predicted, for monodialectal infants, that there would be a significantly smaller proportion of target looks in the 2-feature mispronunciation trials as compared to correctly pronounced trials in the post-naming period, suggesting mispronunciation sensitivity in their native dialect (Mani & Plunkett, 2010; Von Holzen & Bergmann, 2021).

[4] All of our dependent measures are baseline-corrected, to account for potential differences in intrinsic salience across visual items. We obtained baseline-corrected proportion of target looking by subtracting the proportion of time that infants look at the target during the pre-naming window from the proportion of looking time at the target during the post-naming window (see Dependent Measure section below). For the sake of readability, we refer from now on to "proportion of target looking in the post-naming period", omitting the baseline correction.

H2b. We predicted, for bidialectal infants, that there would be no significant difference in the proportion of target looks in the 2-feature mispronunciation trials as compared to correctly pronounced trials in the post-naming period.

Methods

For the current study, we adopted a Sequential Bayes Factor with a maximal sample size approach (SBF + maxN; Schönbrodt & Wagenmakers, 2018). This approach quantifies evidence for both H1 and H0 and requires no adjustments for sampling plans (Wagenmakers et al., 2016). Furthermore, this approach allows for the design to be balanced "in a way that compelling evidence is a likely outcome of the to-be-conducted study, misleading evidence is an unlikely outcome, and sample sizes are within practical limits" (Schönbrodt & Wagenmakers, 2018). In our SBF + maxN design, our maximal sample size will be set to n=60 for each group (monodialectal & bidialectal).

The Bayes Factor (BF) is a factor that uses relative predictive performance, updated as data points are added, to evaluate how likely the hypothesis that an effect is present (H1), or the hypothesis that the effect is absent (H0), is. In other words, it is a "cyclical process of updating knowledge in response to prediction errors…", where "…hypotheses that predicted the data relatively well receive a boost in credibility, whereas hypotheses that predicted the data relatively poorly suffer a decline" (Wagenmakers et al., 2018, p. 1). Following Mani et al. (2021), who implemented an SBF design on three developmental studies, including one on mispronunciation sensitivity, we have first collected data from an initial sample of 20 participants (for each group), and then performing Bayes Factor analyses. Participants will then be tested and added incrementally, one at a time, until the BF exceeds a value of BF=5 (supporting that H1 is more likely) or falls under a value of BF=⅕ (supporting that H0 is more likely), or the maximal sample size for each group (n=60) is met. Given the limits of the thesis timeframe we stopped our data collection at 46 participants but, as needed, more participants will be tested in order to reach the inference criteria outlined in the Stage 1 Registered Report for both groups.

We first applied this procedure on matching trials, to assess whether infants displayed evidence of word comprehension through higher baseline-corrected proportion of target looking times in the post-naming window (cf Dependent Measure section) for hypotheses H1a and H1b. Following this, we continued to incrementally add participants until the inference criteria was reached (BF>5, meaning that it is more likely that infants comprehend words than not, or BF<⅕, meaning that it is more likely that infants do not comprehend words), in each group separately.

When BF>5 in hypothesis H1a (resp. H1b), meaning that it is more likely that infants comprehend words than not, we proceeded further and tested hypotheses H2a (resp. H2b). These hypotheses, H2a and H2b, test whether infants detect mispronunciations, using the full sample used in the former analysis, on H1a (resp. H1b). Note that as the inference criteria for H2a (resp. H2b) was not reached with the existing sample by the time of submission of the present thesis at 26 monodialectal participants, we will continue to recruit and collect data from further monodialectal participants until inference criteria are met. After the first 20 participants in each group were tested, we then performed an SBF analysis after each child on the baseline-corrected proportion of looking time in the post-naming window (cf. Methods, below), comparing correct and mispronounced trials, until the inference criteria were reached (BF>5, meaning that the hypothesis that infants detect mispronunciations or more likely than not, or BF<⅕, meaning that the hypothesis that infants do not detect mispronunciations is more likely) for H2b). In the case that the result of matching trials, to test for word comprehension (hypotheses H1a, resp. H1b) reached the inference criteria BF<⅕ (supporting the hypothesis that infants do not comprehend words), we would stop testing in this group and would not proceed further to test the mispronunciation sensitivity effect (H2a, resp. H2b). The bidialectal group (H1b) reached BF>5 at 20 participants so we went on to further test H2b, however, if the monodialectal group reaches BF<⅕ we will not proceed further to test H2a.

Therefore, in line with the above-mentioned analysis method, 40 infants around 13 months of age (+/- 15 days), living in the Oslo area, were first tested for the study from the National Registry (Folkeregister). The aim for participation in the study is that half (n=20) come from monodialectal families where both parents speak the Eastern (e.g., Oslo) dialect, and the other half from bidialectal Norwegian families, where one parent speaks the local Eastern (Oslo) dialect and the other parent speaks one of the three remaining groups of dialects, i.e., Western, Central or Northern Norwegian. The inclusion criterion in the bidialectal group was a minimum 30% of exposure to each dialect from parental input as assessed by a dialectal exposure questionnaire filled in by parents before their visit to the lab (see below for details). To achieve this aim, at first, parents of 400 age-matching infants were contacted in two waves to take part in the study with their child in each group, which was expected to be sufficient to reach our desired starting sample size (with a 10% response rate and an approximate 50/50 split between mono- and bidialectal households in the Oslo region, as previously observed for similar studies run in the lab). From these outreach mailings we received a high response rate (>10%), but with a higher proportion of monodialectal as compared to bidialectal participants. Even so, we reached a sufficient amount of bidialectal participants to run the first sample set for the Bayesian analysis.

The following criteria have been used to include infants in the study: (1) the child was born full term (gestational weeks >37); (2) the child is exposed to 90% Norwegian or more at home; (3) both parents speak Norwegian natively to the child; and (4) the child has no developmental delays and no history of chronic ear infections.

This study has received ethical approval by the ethics board of the University of Oslo's institute for Psychology.

Stimuli

The stimuli were prepared following the methods used in Mani and Plunkett's (2010) study on onset consonant mispronunciation. We have selected eight familiar (known on average by 57% 13-mo infants, cf. Table 1) consonant-vowel-consonant (CVC) words taken from the Norwegian CDI (Simonsen et al., 2014): *bil, kopp, bok, fot, munn, sokk, ball, vogn;* translated as car, cup, book, foot, mouth, sock, ball, stroller (cf. Figure 1). These words were chosen based on being imageable, monosyllabic, representing diverse onset phonemes, pronounced similarly across the dialects, and having the same (CVC) structure, with the exception of *vogn* which has a CVCC structure. There are two stimuli conditions: correct (e.g., *bil*, [bi:l], car), and two-feature change mispronunciation created by manipulating the onset of the target word (e.g., [ti:l]). All two-feature mispronunciations were manipulated using a combination of voice, place, manner, and/or nasality. We asked native Norwegian speakers, experienced with experimental stimuli in the lab, to assess the stimuli to ensure that the mispronunciations were phonetically and phonotactically legal 'words' in Norwegian. This resulted in eight items for each of the two conditions, or sixteen different trials (cf. Table 1). This set of sixteen trials will be presented twice in a pseudo-randomized order for a total of thirty-two trials, and we have ensured that the same stimuli will not be presented twice consecutively. In addition, we created two stimuli to be used in the outcome neutral condition: *hund* (dog) and *katt* (cat). In these trials, presented once at the beginning of the experiment and once halfway through the experiment, the images are presented in isolation and counterbalanced on either the left or right side of the screen along with matching audio. Looking times from these trials have been used to exclude non-engaged participants, i.e., those who do not direct their gaze toward the stimuli for at least 50% of each trial in response to their presentation.

Word *(translation)*	Transcription	CDI 13mos	Cohort size 50%	Cohort size 10%	2-feat. Δ	Transcription	Cohort size 50%	Cohort size 10%	*Paired image*
bil *(car)*	bi:l	75	41	14	t	ti:l	16	4	munn
kopp *(cup)*	kɔp:	52	16	4	b	bɔp:	41	14	fot
bok *(book)*	bu:k	64	41	14	j***	ju:k	7	1	vogn
fot *(foot)*	fu:t	42	16	0	l***	lu:t	14	2	kopp
munn *(mouth)*	mʉn:	47	15	5	d****	dʉn:	15	3	bil
sokk *(sock)*	sɔk:	43	37	6	v***	vɔk:	13	2	ball
ball *(ball)*	bal:	84	41	14	l**	lal:	14	2	sokk
vogn *(stroller)*	vɔŋn	48	14	3	n	nɔŋn	8	3	bok
Mean (SD)		57 (15.72)	27.63 (13.31)	7.5 (5.66)			16 (10.61)	3.88 (4.20)	

voicing & place

**place of articulation
***manner and place
****manner and nasality

Table 1. Audio stimuli for trials, including transcriptions, cohort sizes at 50% and 10%, and CDI reported comprehension at 13 months.

Figure 1. Visual stimuli for target word and outcome neutral trials.

Cohort sizes of all onset consonants presented in the trials were computed from the Norwegian CDI both at the 10% and 50% threshold (i.e., the number of given words starting with that sound that 10 (resp. 50) percent or more of parents reported their child understanding at 13 months of age). We also conducted a Bayesian analysis in R with the BayesFactor package (Morey et al., 2021) to examine the likelihood of the average cohort sizes of each stimuli condition (correct and 2-feature change) being different from the null and found no significant effects (10%: BF = 1.43. 50%: BF = 0.85), eliminating a possible confounding effect due to an imbalance in cohort sizes across stimuli conditions (cf. Table 1).

A native Norwegian speaker, who was born and grew up in Oslo (Eastern dialect), was recorded while reading in a child-directed fashion the carrier sentence: 'Look! <target>!'. The target words are the eight labels of the items depicted on the pictures, each in the two conditions: correct word and 2-feature change. The outcome neutral item labels have also been recorded in the same fashion. Recordings were done in pairs to ensure similar tones, F0 contours and tempo between the correctly and mispronounced words. The following parameters were used for recordings: 16 bits, 2 channels, 22.05 kHz.

The recordings were checked for auditory quality and saved to individual .wav files. When necessary, some residual noise in the recordings was removed in Praat (Boersma & Weenink, 2007) using the *remove noise* function (window length 0.025 s, filter 80 – 10 kHz, smoothing bandwidth 40 Hz, and spectral subtraction noise reduction method). All eighteen word types were also equated for amplitude: the mean amplitude will be set to 65 dB. The resulting audio files (described above) will be used to prompt infants' looks at the images presented in the IPL task.

Ten depicted items have been used as visual stimuli. Eight of them depict objects referred to with the target words used in the study, with two exemplars for each item (e.g., a blue car and a red car), and two images, a cat and a dog, were used for the outcome neutral condition (see Figure 1). The imaged objects were also edited so that their relative brightness and size are approximately the same. The eight picture pairs, assembled from two versions of the target word images, are laid out on a light-grey background 51 cm by 28 cm, i.e., the size of the experimental screen used for the study. To counterbalance the order of presentation, corresponding audio files (n=16) with their picture pairs (n=8) are presented in pseudo-random order and the target image is presented equally often on the left and right side of the screen. Between the trials, if the infant looks away, a spinning wheel along with a jingling noise will be presented to regain the infant's attention (see Procedure for details).

Dialectal exposure questionnaire

Infants' exposure to each Norwegian dialect at home was collected from the dialectal exposure questionnaire that primary caregivers (although we will use "parents" throughout the manuscript) fill in prior to their lab visit. To measure home parental dialectal exposure, we ask parents to indicate how much time (in %) the child hears one parent's speech in relation to the other parent's speech. The sum of parental input should equal 100%. Previous studies have found that the effects of the home environment have the most impact, above differences in language exposure in other environments (Bosch & Ramon-Casas, 2011; 2009).

Word exposure questionnaire

In addition, parents fill out a word exposure questionnaire. This questionnaire asks, for each of the eight words used in the task, how frequently parents would have used it (on a scale from 0-never to 5-very frequently) while interacting with the child or in her presence, since their baby was born. Words reported as being not heard by the child will be excluded from the analyses.

Word knowledge (CDI) questionnaire

Parents also fill out the Norwegian version of the McArthur-Bates Communicative Development Inventory (CDI) (Simonsen et al., 2014) to provide a measure of receptive vocabulary size, which will be used in exploratory analyses looking into potential associations between vocabulary size and mispronunciation sensitivity. We will report CDI sizes and group comparisons using a Bayesian t-test after data collection, though vocabulary sizes of 12- and 24-month-olds, as assessed with Norwegian CDIs, were found not to differ between children raised in monodialectal families and those raised in bidialectal families (Munoz, Mayor & Kartushina, in prep). Currently, the Norwegian CDI does not have dialectal variations in its inventory; none of the stimuli used in this study have explicit dialectal differences.

Procedure

Data collection was performed in single sessions at the BabyLing laboratory, at the University of Oslo's institute for Psychology, equipped with an EyeLink 1000 Plus eye-tracker. Prior to the lab visit, parents received an information letter briefly presenting the aims of the study and filled in the questionnaires (see above), i.e., a dialectal exposure, word exposure, and word knowledge questionnaire using an online platform, Nettskjema, provided for academic use by the University of Oslo's institute for Psychology (similar to Qualtrics). A native speaker of Norwegian received the parent(s) with their child in the reception room of the lab and then briefly explained the task that the child would perform. Then the parent, having signed the consent form before coming to the lab, and the experimenter accompanied her to the eye-tracking room with her child.

The Intermodal Preferential Looking paradigm task was performed using an EyeLink 1000 Plus eye-tracker, placed on the EyeLink LCD Arm Mount and using a 500 Hz sampling rate (monocular), infant mode calibration and a 1280 x 1024 pixels screen resolution. The child sat on her parent's lap facing the experimental computer screen fitted with an eye-tracker base. Parents also wore sound-attenuating headphones through which they heard masking music (a custom blend of instrumental music and a pastiche of randomly timed and random amplitude stimulus materials from the ManyBabies Consortium, 2020). The parents are asked not to talk to the child, point to the screen or shift their bodies. The experimenter always sat in the same room, behind the parent, so neither the child nor the parent would see her during the task. The stimuli were all presented to the child at the average amplitude of 65 dB through two speakers, integrated in the left and right sides of the monitor. The experimenter was also able to monitor infants' looking behavior via the EyeLink Live Viewer tool operating on the control screen, situated in front of the experimenter and to ensure that the child did not become distressed.

The experiment started with an automated 5-point calibration procedure (infant version), followed by a validation procedure, then an outcome neutral trial was presented to start the task. This trial consists of a single image (*dog* or *cat*, presented randomly on one side of the screen) presented with the correct label ("dog" or "cat") to gauge attention to the screen. Then the mispronunciation task with the picture pairs began. The order of the stimuli in the task has been counterbalanced across participants, and we ensured that there was always at least one trial between any correctly pronounced and mispronounced trials for the same word. The second outcome neutral trial was presented halfway through, after presentation of the sixteenth trial, to check for continued attention. All infants were presented with all stimuli and conditions in one of four lists which has counterbalanced the item presentation order, unless the experiment was terminated for any reason (cf. Exclusion Criteria, below).

Each trial (thirty-four in total, two of which were outcome neutral) started with the presentation of a small fixation cross on a grey background. If the child fixated the cross continuously for 400ms, the experiment continued automatically to the stimuli presentation. If the child did not fixate the cross, for at least 400 ms, within 2000 ms, a colorful spinning wheel appeared in the middle of the screen, accompanied by the sound of jingling and a bird tweeting, to attract the child's attention. If the child then fixated on the spinning wheel continuously for 500 ms, the experiment continued to the stimuli presentation. If the infant did not fixate on the spinning wheel within 90 sec, the experiment would be aborted. For the stimuli presentation, a picture pair (see Stimuli section above) appeared on the screen for 1.5 sec before the paired auditory stimulus was presented (one of two conditions for each picture: word or 2-feature change). The picture remained on the screen for 3.5 sec after the target word onset. The task stopped after all thirty-four trials were presented to the infant.

At the end of the experiment parents were able to choose a small gift for their infant (e.g., a toy) and reimbursed for travel costs.

Data preprocessing

Exclusion criteria

The following criteria have been used to exclude infants from data analysis based on their behavior: (a) failed calibration of the eye-tracker; (b) software problem (e.g., technical reasons: software stops displaying images or playing sounds for more than 50% of the trials); (c) the parent withdraws consent, or either parent or child show signs of discomfort; (d) the child did not contribute to at least six valid trials in each condition for the experimental data, that is, a minimum of twelve experimental trials in total.

Single trials in the mispronunciation task were excluded if there was less than 500ms of recorded looking in the pre-naming period and/or no recorded looking at either picture for at least 50% of the post-naming period. Individual item trials were removed from individual child data if parents reported in the word exposure questionnaire that their child had not heard the word. Finally, we excluded trials in which the experimenter reported that the parent interfered (e.g., pointed to the screen, shifted his/her body, or moved his/her chair), or the trial was interrupted by a third person or due to a technical error. If the experimenter heard crying, then she terminated the experiment.

Interest areas and window of analysis

Interest Areas (IAs), left and right, were determined prior to data collection and based on the image dimensions plus 165 pixels on the top and bottom of each image. IAs were the same for all trials. Looks captured with the EyeLink eye tracker in each interest area were used to calculate the outcome measure. Our outcome measure is the baseline-corrected proportion of target looking, computed by subtracting the proportion of time that infants looked at the target during the pre-naming window from the proportion of looking time at the target during the post-naming window. The pre-naming window is the start of the trial until the target word onset, and the post-naming window of analysis is set to 700-2500ms after target word onset. This was calculated from previous studies with monodialectal infants (Adams et al., 2018; Garrison et al., 2020) utilizing PeekBank (Zettersen et al., 2021). See below for a justification of the analysis window.

$$\frac{postnaming\ T}{(postnaming\ T + postnaming\ D)} - \frac{prenaming\ T}{(prenaming\ T + prenaming\ D)} \qquad T = Target, D = Distractor$$

Therefore, the outcome measure (Prop_target) varies between -1 and 1; where 0 indicates an absence of preference for either picture. The Sequential Bayes Factor (SBF) approach does not allow for a data-driven approach to identify an optimal window for analysis using our own data (since a flexible analysis window would impact evidence for H0 and H1). However, to optimize our pre-registered window of analysis, we examined existing eye-tracking studies that have adopted a similar question and design (Adams et al., 2018; Garrison et al., 2020) and used PeekBank (Zettersen et al., 2021) to calculate the optimal post-naming time window of analysis, for infants of matching ages (12-14m, n=16). The result from this analysis suggests the optimal post-naming analysis window should be from 700 to 2500ms post-target-word-onset (see Fig. 2).

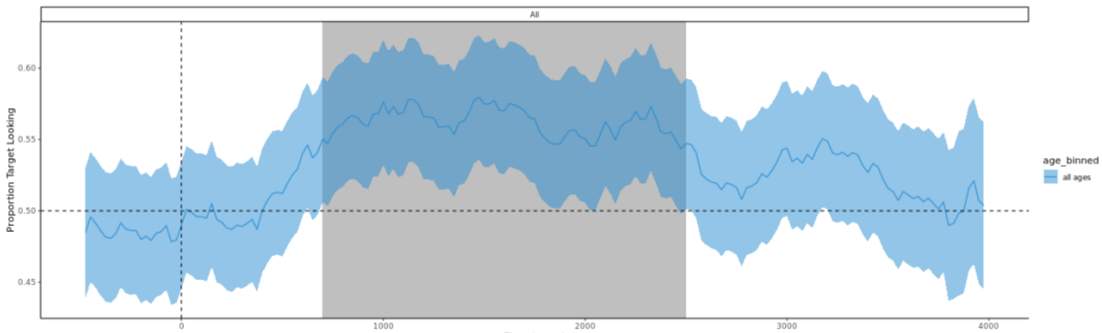

Figure 2. Optimal post-naming time window of analysis calculated using PeekBank.

Results

After applying the participant-based exclusion criteria, eight infants were excluded, and, upon application of trial-based level exclusion criteria, 49 trials were excluded (resulting in a remaining 96% of the total data). The minimum number of trials was fourteen per child: six correct trials, six mispronounced trials, and one of the outcome neutral trials. We then performed analyses on the remaining valid participants and trials. All presented analyses were performed using sequential hypothesis testing with Sequential Bayes Factors (Schönbrodt et al., 2017), in line with Mani et al.'s (2021) approach with developmental studies. These SBF analyses thus assess whether 13-month-old monodialectal and bidialectal Norwegian infants comprehend correctly pronounced familiar words, and whether they detect onset consonant mispronunciations of these familiar words.

Group	**Number**	**Boys/girls**	**Age in days mean (SD)**
Monodialectal	26	14/12	397.54 (7.37)
Bidialectal	20	8/12	401.45 (9.93)

Table 2. Final sample of participants after exclusion criteria were implemented.

Word comprehension (H1a, H1b)

First, Bayesian t-tests (which tell us how likely H1 – that there is an effect – is, compared to H0 – that there is no effect), as implemented in the BayesFactor package for R (Morey et al., 2021; R Core Team, 2012), were performed separately for each group of infants (n=26 monodialectal, n=20 bidialectal) in the correctly pronounced trials, to test whether infants displayed evidence of word comprehension (hypotheses H1a and H1b). The dependent variable was the baseline-corrected PTL of monodialectal and bidialectal infants, on correctly pronounced valid trials (as described above).

As shown in Figure 3, in line with H1a, monodialectals exhibited a small yet non-significant effect of naming[5], providing anecdotal evidence for support of word comprehension with a greater proportion of looking times at the target image in the post-naming phase as compared to the pre-naming phase (M = 0.041, SD = 0.098, BF = 1.401). In line with H1b, Bidialectal infants exhibited moderate evidence of naming according to Bayesian standards (M = 0.068, SD = 0.098, BF = 8.143). Both results are interpreted following Aczel et al.'s (2017) Bayes factor levels of evidence (anecdotal, moderate, strong).

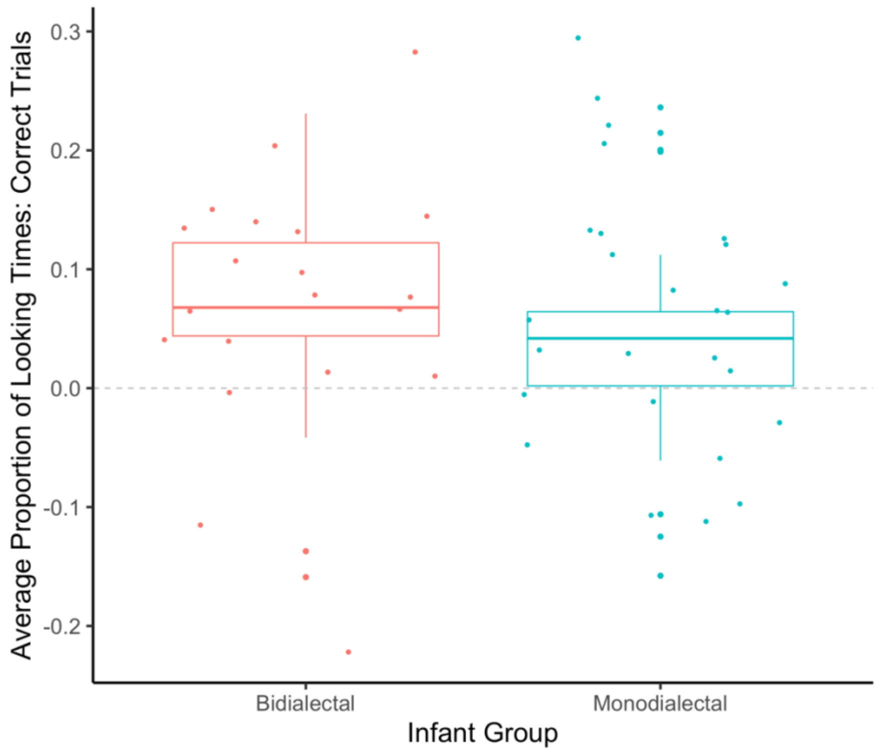

Figure 3. Average proportion of looking times for correct trials by group. The horizontal dotted line represents a PTL of 0, or no preference for either image post-naming.

We also conducted an exploratory Bayesian t-test to further examine the difference between the two groups' response to correct trials. The resulting Bayes Factor revealed anecdotal evidence for the absence of a difference, that is, monodialectals and bidialectals do not differ

[5] It is worth noting that although evidence was anecdotal by Bayesian standards (BF = 1.04), a standard t-test resulted in a significant p-value of .044. This observation, explored in-depth by Aczel et al. (2017), suggests that a frequentist approach (reporting p-values), more traditionally used in psychological research, may not be as compelling as previously accepted, and is worth considering when analyzing study results. See the Discussion section for further commentary.

significantly in response to correctly pronounced words (BF = 0.424). This suggests that the two groups are much more likely to be similar to, than different from, each other in word comprehension.

Even so, presently, only Hypothesis H1b meets the inference criteria, with a BF > 5, suggesting that 13-month-old infants raised in bidialectal families displayed evidence of word comprehension. However, Hypothesis H1a has not reached a BF of > 5 or < ⅕. Thus, as of now with the present sample, we cannot conclusively determine whether monodialectal 13-month-old infants do or do not display word comprehension for the correctly pronounced trials presented. Figure 4 (below) displays how the BF, for monodialectal infants, has changed as evidence has been added thus far. However, as mentioned above and following the Registered Report, participants will be added incrementally while the MA thesis is being evaluated until the inference criteria, BF > 5 or BF < ⅕, is met.

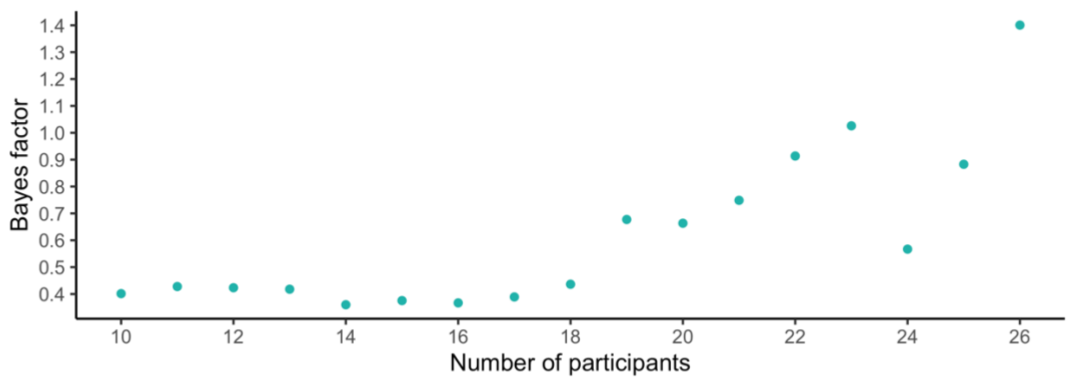

Figure 4. Resulting Bayes factor for monodialectal infants tested on word comprehension, as participants were added, starting at 10 participants.

Mispronunciation sensitivity (H2a, H2b)

Although there is no conclusive evidence that monodialectal infants either do or do not display word comprehension (H1a), we have chosen to report, for the sake of consistency, the preliminary results from the monodialectal group at the time of the thesis submission below. So, for both infant groups separately, we then addressed our second hypothesis, whether infants displayed evidence of sensitivity to mispronunciations, by examining the baseline-corrected PTL of bidialectal infants on the 2-feature mispronunciation trials as compared to the PTL on the corresponding matching trials.

Monodialectals exhibited moderate evidence (BF = 0.105) to support H0, or a lack of mispronunciation sensitivity, with only a slightly greater proportion of looking times at the target image in the post-naming phase for correct trials (M = 0.041, SD = .098) as compared to mispronounced trials (M = .021, SD = .093). However, as we can see in figure 5a (below), the distributions between the conditions were markedly different. Monodialectal infants displayed a large variation in the PTLs for correctly pronounced trials with a high concentration of participants with average PTLs around the mean (0.041), whereas the mispronounced trials yielded more spread out PTLs with a slight congregation around the mean (0.021).

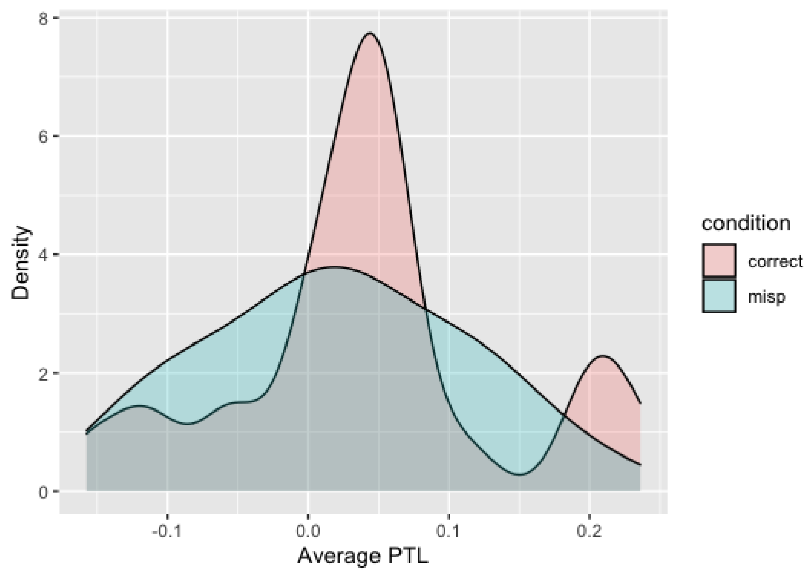

Figure 5a. Average proportion of looking times across trials for monodialectal infants by condition.

As shown in Figure 5b, bidialectal infants exhibited moderate evidence (BF = 10.214), by Bayesian standards (Aczel et al., 2017), of mispronunciation sensitivity, with a greater proportion of looking times at the target image in the post-naming phase for correct trials (M = .068, SD = .098) as compared to mispronounced trials (M = -.021, SD = .086). The Bayes Factor with the existing sample size has reached the upper inference criteria, of BF > 5, providing evidence that bidialectal infants display mispronunciation sensitivity at 13 months of age. Furthermore, below (see Fig. 5b) we can observe two distinct distributions of PTLs for the correctly pronounced and mispronounced trials separately.

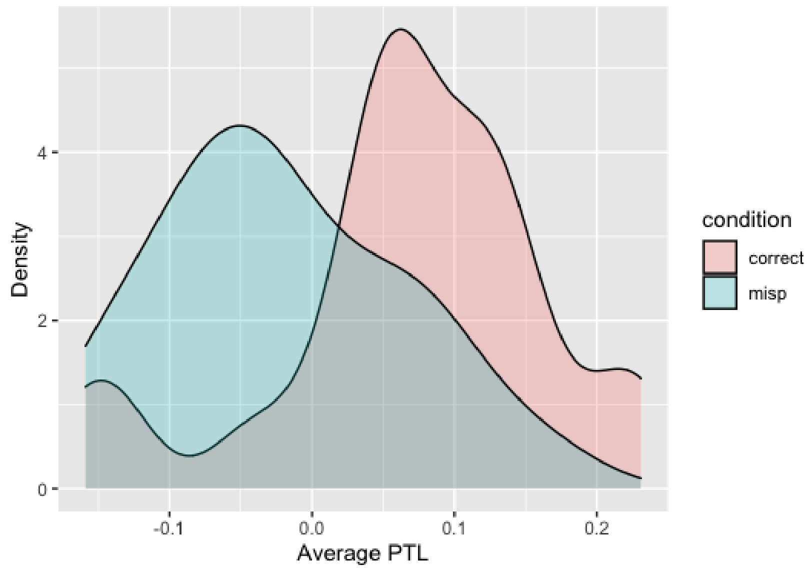

Figure 5b. Preliminary results; average proportion of looking times across trials for bidialectal infants by condition.

Upon further examination of individual stimuli trials for both infant groups, we also see that there was wide variability in the response to the individual trials. Although a clear difference between the correct and mispronounced trials can still be observed overall for the bidialectal infants (Figure 6b), monodialectal infants do not display a difference between correct and mispronounced trials to the same degree (Figure 6a).

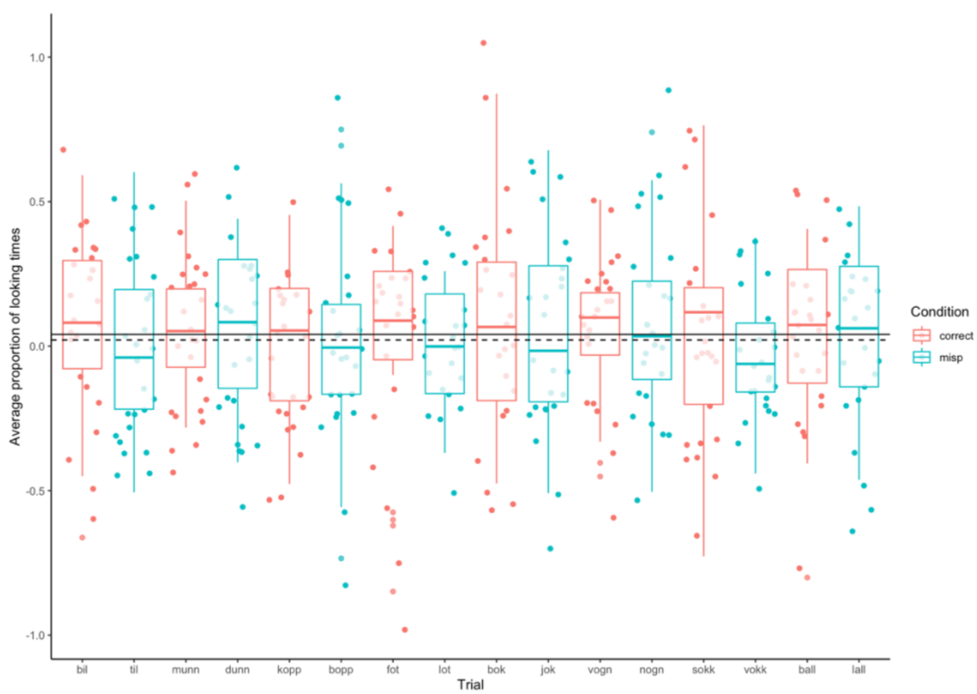

Figure 6a. Average PTL of monodialectal infants by trial, correct and mispronounced trials side-by-side. The solid line represents the average PTL for correct trials (in red), while the dashed line represents the average PTL for mispronounced trials (in blue).

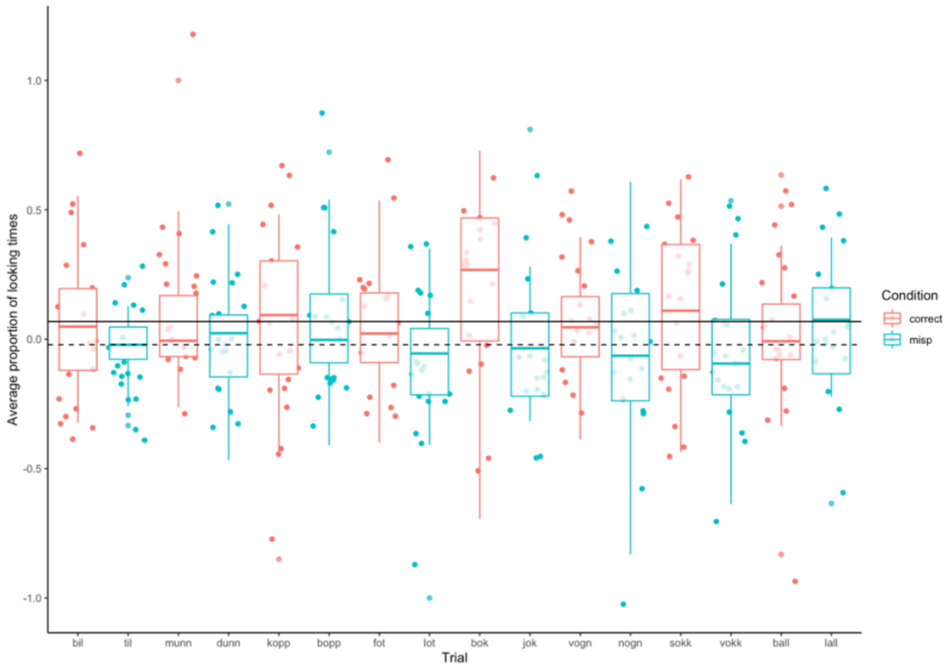

Figure 6b. Average PTL of bidialectal infants by trial, correct and mispronounced trials side-by-side. The solid line represents the average PTL for correct trials (in red), while the dashed line represents the average PTL for mispronounced trials (in blue).

Discussion

The present study was inspired by previous research on mispronunciation sensitivity in monodialectal infants (Mani & Plunkett, 2010), as well as dialectal variability in older toddlers (Durrant et al., 2015; Floccia et al., 2012) to explore how infants deal with native dialectal variation around their first year of life. To examine whether exposure to dialectal variability from birth, as indicated by PTL measures in response to presented stimuli, plays a role, we used a mispronunciation paradigm in an IPL task and examined the looking times of both mono- and bidialectal age-matched participants in response to presented target words that were either correctly pronounced or mispronounced (e.g., *vogn* and *nogn*). The mispronunciation paradigm is a way to probe for mispronunciation sensitivity, where a difference in looking times at the target when paired with correctly pronounced versus mispronounced label indicates that the infant registers the difference and is thus sensitive to the mispronunciation.

Word comprehension

We first analyzed the looking times of both monodialectal and bidialectal infants in response to the correctly pronounced stimuli to establish word comprehension in each group. Hypotheses H1a and H1b targeted word comprehension by examining baseline-corrected proportion of target looks (PTL) in the post-naming period. A baseline-corrected PTL greater than 0 was taken as evidence that the infant looked longer at the target image upon being named and was thus able to recognize the relationship between the target image and its corresponding label.

Results revealed that generally, infants looked slightly longer at the target in the post-naming period in correctly pronounced trials. Monodialectal infants displayed anecdotal evidence for an effect of naming (BF = 1.401), whereas bidialectal infants showed a strong effect of naming (BF = 10.214), also meeting the inference criteria. There was a marked difference between the two infant groups, with bidialectal infants exhibiting strong evidence of word comprehension at 13 months of age whereas monodialectal infants did not reach decisive evidence supporting word comprehension with the current sample. This result is partly at odds with recent research by Kartushina and Mayor (2022), where word comprehension was observed in monodialectal, but not bidialectal, Norwegian 12-month-old infants. While it may be argued that the additional month of language input (here, with 13-month-old infants) allowed Norwegian bidialectal infants to catch up to their monodialectal peers on word comprehension, this would not explain why monodialectal infants may display evidence of word comprehension at 12 months of age, but not at 13 months of age.

There are some possible interpretations of the differences in the effect between monodialectal and bidialectal infants on word comprehension. One reason for this discrepancy could be the method of analysis (first mentioned in Results, footnote 5). Aczel et al. (2017) explored the relationship between frequentist (reporting p-values) and Bayesian (reporting Bayes factors) approaches in psychological research and argues that although p-values often correspond with Bayes factors in terms of the support for which direction the hypothesis leans, p-values "systematically overestimate the strength of the evidence against the null-hypothesis" (Aczel et al., 2017, p.3). This suggests that a more traditional analysis of word comprehension data may be less stringent, producing significant results in the previous study by Kartushina and Mayor (2022) as well as others. Furthermore, Dienes (2014) also argues that non-significant results are still valuable for assessing the degree to which a hypothesis is supported by the evidence in a study. In our case, while the monodialectal word comprehension hypothesis resulted in BF of 1.401 with the current participant group (n=26), it is still anecdotal evidence, and further participant testing to reach the inference criteria will allow us to make conclusions about monodialectal word comprehension with more certainty.

Further, our study did not find a significant difference between groups, which is in line with Kartushina and Mayor's (2022) study with 12-month-old participants. It is currently unclear what drives this difference in PTLs in each group, if age or choice of stimuli play a role. It is also worth noting that participants in both monodialectal and bidialectal groups in the current study exhibited a large variation in PTLs. This suggests that despite variability, there is a significant group effect of naming already in the early stages of language learning.

The trial stimuli were carefully considered and chosen because they have been reported through the CDI (Simonsen et al., 2014) as understood by a high proportion of 13-month-olds (see Table 1). Even so, they are different from the study by Kartushina and Mayor (2022), which might also contribute to the current study's results differing from previous research. Also, the current study had a total eight images with two depictions of each, whereas Kartushina and Mayor (2022) had sixteen separate images. Further analysis may be conducted to explore if there was a difference in response between the visual stimuli versions. All of these factors, in addition to those mentioned above in the Word comprehension section, may potentially account for some differences in study results.

Mispronunciation sensitivity

We then compared the looking times of infants in response to correct and mispronounced stimuli to determine mispronunciation sensitivity. In line with the Registered Report, we only evaluated mispronunciation sensitivity (Hypotheses 2a and 2b) if comprehension was established in Hypotheses 1a and 1b. Thus, we restricted the analysis of mispronunciation sensitivity to the bidialectal group. However, as mentioned in the Results section, we did choose to explore and report preliminary results from hypothesis H2a with participant data collected at the time of thesis submission.

Hypothesis H2b targeted mispronunciation by examining baseline-corrected proportion of target looks (PTL) in the post-naming period for correct compared to mispronounced trials. Bayesian t-tests revealed that bidialectal 13-month-old infants exhibited strong evidence of mispronunciation sensitivity, with a lower PTL in the post-naming period in mispronounced trials compared to correct trials. This finding in the bidialectal group is in line with previous IPL studies on consonant mispronunciation sensitivity in monodialectal infants at 12 (Mani & Plunkett, 2010) and 14 (Ballem & Plunkett, 2005) months, as well as 20-month-old bidialectal infants (Durrant et al., 2015).

Surprisingly, bidialectal infants displayed strong sensitivity to onset consonant mispronunciations, suggesting that they may take more time to evaluate whether the distractor was perhaps a potential target rather than a match between the mispronounced target word and the target image. This contrasts with Durrant et al.'s (2015) study findings where bidialectal infants displayed less specific lexical representations at the time of the study, suggesting that from-birth variable input resulted in generally more relaxed lexical boundaries. However, as studies by Schmale et al. (2011) and White and Aslin (2011) posit, the flexibility, or relaxation, of phonetic boundaries by bidialectals is input-specific and thus might not apply to the mispronounced trials presented in the study. Additionally, our bidialectal sample consisted of twelve girls and eight boys, which may also have influenced the results. Research by Eriksson et al. (2012) found that girls were generally ahead of boys in emerging language skills and thus, our sample heavily skewed by girls could have resulted in a higher Bayes Factor for the bidialectal group.

Upon preliminary analysis of H2a, we found that the monodialectal infants in this study responded distinctly from their bidialectal peers, exhibiting similar looking times at the target in the post-naming period whether the target label was mispronounced or not. This is at odds with research by Mani and Plunkett (2010) and Ballem and Plunkett (2005), who detected consonant mispronunciation sensitivity of familiar words in monodialectal infants at 12 and 14 months of age, respectively. However, in a recent mispronunciation sensitivity meta-analysis, Von Holzen and Bergmann (2018) suggested that even though the infants looked at the target regardless of pronunciation, they may more systematically look at the target when correctly pronounced. Another explanation could be that although the infants recognize that the word is mispronounced, they still accept the label as more suitable to the target image in comparison to the distractor.

Limitations

Dialectal exposure

Dialectal exposure in Norway is especially common, and although we have controlled for parental input, it is important to note that these infants are not growing up in completely isolated linguistic environments. Input from other caregivers is inevitable, as well as from possible television or radio programming. Additionally, due to the unique diversity of Norwegian dialects, many extended family members (such as grandparents or aunts and uncles) can speak dialects different from the parents.

The similarity of dialects the infant is exposed to may also play a role in mispronunciation sensitivity. Kartushina and Mayor (2022), in their recent examination on dialectal similarity and exposure in Norwegian 12-month-old infants, found that the reported similarity in parental dialects of bidialectal infants played a role in early word representations, where bidialectal infants' word comprehension task performance was negatively affected as the dissimilarity between parental dialects increased. This suggests that the extent of dialectal variability that children are exposed to likely contributes to less specified phonological representations.

Our study also collected information from parental dialectal input, yet it was limited to the main dialect categories: Eastern (østnorsk), Western (vestlandsk), Central (trøndersk), and Northern (nordnorsk). Although the perceived dialectal variability was not reported, further analysis may reveal some effect between dialects spoken at home on word comprehension. Additionally, during their visit to the lab, after the experiment several parents expressed that they felt they had a "watered out" ("*utvannet*") dialect from living in Oslo, dropping the use of dialect-specific words or pronunciations, or speaking with less tonal inflections. However, after visiting their home region or in speaking with others that had the same dialect, their dialect would be more noticeable. While this was not formally reported, it is an interesting aspect to take into consideration, where individual speakers may also present variation in pronunciations.

Mispronunciation type and location

In the current study we tested consonant onset mispronunciation only, and although dialect-specific onset consonant variability exists in Norwegian dialects (such as /v/ to /k/ in some words, see Introduction), Norwegian dialects also feature medial consonant changes (such as /t/ to /d/ in "båt", *boat*, in addition to medial vowel changes such as /o/ to /u/ in "tror", *thinks*.

Testing medial vowel mispronunciations as well, such as in previous work by Mani and Plunkett (2010), may give more comprehensive insight into mispronunciation sensitivity around this age, as well as eventual dialect-specific phonological flexibility.

Intermodal Preferential Looking (IPL) design

The current study, along with many others, have used IPL tasks in the field of lexical development. However, this method can present challenges when determining sensitivity to mispronunciation and calculating looking times. Individual participants may have a preference for selected stimuli in the trials, and a recent meta-analysis on infant mispronunciation response by von Holzen and Bergmann (2021) revealed that trial time differences across studies can impact calculations and interpretations of looking times. Moreover, potential overlap of the onset consonant of words associated with target and distractor images produces greater mispronunciation sensitivity than those whose labels do not overlap. Furthermore, Mayor and Plunkett (2014) applied the TRACE model of word recognition, an incremental process that eliminates competing candidates in one's mental lexicon (McClelland & Elman, 1986), to model speech perception in infancy and early childhood and provided a mechanistic interpretation of the why distractor images can confound the interpretation of looking preference.

Alternatively, pupillometry may provide the opportunity to examine the cognitive activity related to presented stimuli in an experiment while eliminating (potentially confounding) distractors. Pupil dilation has been shown to index increased cognitive load, surprise, and arousal, including among children (Beatty & Lucero-Wagoner, 2000; Karatekin, 2007), making pupillometry a useful tool to capture a range of phenomena associated with language processing. Within the context of language development and related to the current study, recent studies have demonstrated that pupillometry can be used to study children's sensitivity to mispronunciations as they observed greater pupil dilation in response to mispronounced versus to correctly pronounced labels (Fritzsche & Höhle, 2015; Tamasi, 2016; Tamasi et al., 2017). Tamasi and colleagues (2019) recently explored the effects of degrees of lexical feature manipulation (i.e., place, manner, and voicing changes) using an IPLP task paired with pupil dilation measurement and found that the featural distance from the correct label of a target item did in fact affect the looking behavior. Although the study was conducted with monodialectal, 30-month-old children, it opened up the possibility of further examination into how subsegmental manipulations of word (mis-)pronunciations can be detected at younger ages, such as 12 to 14 months of age.

Additionally, as the influence of dialectal exposure on the developmental trajectory of lexical representations is still unclear and involves a myriad of segmental and subphonemic changes, pupillometry as the mode of data collection might give a new window through which to closely examine any reactions as the stimuli are presented. A mispronunciation sensitivity study using pupillometry as the method should thus be considered to determine if further mispronunciation studies in younger infants would benefit from a pupillometry design.

Future Research

Cluster permutation analysis

In alignment with the Stage 1 Registered Report, we will perform cluster permutation analyses after complete data collection is conducted to provide insights into the differences (if any) in dynamics of mispronunciation sensitivity in bidialectal as compared to monodialectal infants. These will not be used to make any inferences from a Bayesian perspective, nor to determine a stopping point for data collection, but rather to explore whether the analysis window chosen was the best one. As the existing literature on mispronunciation sensitivity in infants is limited, we determined our window of analysis for this study using PeekBank (Zettersen et al., 2021). Performing the cluster permutation analyses on the current study will allow for a more nuanced examination of the window of analysis and inform similar future studies.

Sequential Bayes Factor (SBF) analysis

The use of Bayesian analysis in developmental research is a new, yet promising approach. Inspired by Mani et al.'s (2021) research replicating existing studies utilizing SBF, we also applied this to the current study. We experienced that the Bayesian statistic approach may be more stringent than the more traditional frequentist statistic, as exhibited in the results of H1a. This is also supported by research by Aczel et al. (2017), who found that just over half of the >50,000 psychology studies he examined the significant p-values of had corresponding Bayes factors with strong compelling evidence for the alternative hypothesis. Even so, utilizing this method of analysis might lead to a push toward reliable and compelling future research contributions.

While we were able to recruit enough participants for the initial sample size within the timeframe of the master thesis, additional participants will be recruited to reach the desired sample size for monodialectal participants, so as to reach the inference criteria. Although more testing is needed, one potential additional benefit of utilizing SBF is testing less participants. For example, Mani et al. (2021) met inference criteria and stopped data collection at 32 participants, as opposed to 55 participants in the original study. In the context of the current study, we met inference criteria at 20 bidialectal participants and have currently tested 26 monodialectal participants, whereas Kartushina and Mayor (2022) collected data from 35 participants in each group. These differences exhibit the efficiency of the SBF approach and end up saving much collective time and resources put into both participant recruitment and data collection.

Conclusion

Many infants grow up in countries with different dialects and are consequently exposed to dialectal variability. Yet, there has been little research on how from-birth dialectal variability may impact early language learning development. This study aimed to explore more specifically how dialectal variability impacts mispronunciation sensitivity around the first year of life. Variable input, here through two parents speaking different native dialects, has shown to play a contributing role in this process. Norway's rich linguistic environment poses a unique opportunity to study native dialectal variations and its effects on language learning; how a language takes shape is a remarkable feat that involves sorting out which speech sounds belong to the same, or different, phonemic categories. As results of this study reveal, discerning these speech sounds takes time, exposure, and consistency. Further, phonological word representations have been shown to be more robust in bidialectal infants compared with their monodialectal peers around 13 months of age.

While there is still much unknown about the process of building lexical representations in infants around 13 months of age with variable dialectal input, this study allowed us to, thus far, establish word comprehension in bidialectal infants, as well as lay groundwork for follow-up studies focused on dialectal exposure and lexical representations using mispronunciation sensitivity. This might also allow us to explore any possible differences in the trajectories of language learning in Norwegian monodialectal and bidialectal infants.

Perhaps the key takeaway of this study at the time of thesis submission is that monodialectal and bidialectal infants do not process words the same, and that only bidialectals show mispronunciation sensitivity. Although these findings were somewhat surprising, while bidialectal infants exhibited strong effects of naming as well as significant sensitivity to mispronunciations, monodialectal infants showed an anecdotal (yet inconclusive by Bayesian standards) effect of naming. Preliminary analyses reveal slight mispronunciation sensitivity, though we cannot say until eventual word comprehension is established and meets the inference criteria (data collection is ongoing).

So, from-birth variability in speech in the form of native dialectal differences does affect lexical specificity in early word learning. Results from the study thus far reveal insight to the specificity of lexical representations that Norwegian monodialectal and bidialectal infants have at 13 months of age and may in turn lead to a better understanding of how these groups of infants build their lexical representations and how specific they are around their first year of life.

Even so, additional work in this field is needed to determine how, and if, differences in lexical representations change over time in the context of mispronunciation sensitivity in bidialectal infants. First, however, testing more participants within this study is needed for more conclusive results in line with the Registered Report and within the Bayesian statistical framework.

References

Aczel, B., Palfi, B., & Szászi, B. (2017). Estimating the evidential value of significant results in psychological science. *PLOS ONE*, *12*, e0182651. https://doi.org/10.1371/journal.pone.0182651

Adams, K. A., Marchman, V. A., Loi, E. C., Ashland, M. D., Fernald, A., & Feldman, H. M. (2018). Caregiver talk and medical risk as predictors of language outcomes in full term and preterm toddlers. Child Development, 89(5), 1674-1690. https://doi.org/10.1111/cdev.12818

Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language*, *32*(1), 159–173. https://doi.org/10.1017/S0305000904006567

Bergelson, E. (2020). The comprehension boost in early word learning: Older infants are better learners. *Child development perspectives*, *14*(3), 142-149.

Bergelson, E., & Aslin, R. (2017). Semantic specificity in one-year-olds' word comprehension. *Language Learning and Development*, *13*(4), 481-501.

Bergelson, E., & Swingley, D. (2018). Young Infants' Word Comprehension Given An Unfamiliar Talker or Altered Pronunciations. *Child Development*, *89*(5), 10. https://doi.org/10.1111/cdev.12888

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of Phonological Constancy. *Psychological Science*, *20*(5), 539–542. https://doi.org/10.1111/j.1467-9280.2009.02327.x

Boersma, P., & Weenink, D. (2007). *PRAAT: Doing phonetics by computer (Version 5.3.51)*.

Bosch, L., & Ramon-Casas, M. (2009). Phonetic variability in bilinguals' acquisition of native-vowel category contrasts. *The Journal of the Acoustical Society of America*, *125*(4), 2770–2770. https://doi.org/10.1121/1.4784720

Bosch, L., & Ramon-Casas, M. (2011). Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics*, *39*(4), 514–526. https://doi.org/10.1016/j.wocn.2011.02.001

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous Bilingualism and the Perception of a Language-Specific Vowel Contrast in the First Year of Life. *Language and Speech*, *46*(2–3), 217–243. https://doi.org/10.1177/00238309030460020801

Cerruti, M., & Tsiplakou, S. (2020). *Intermediate Language Varieties: Koinai and regional standards in Europe*. John Benjamins Publishing Company.

Curtin, S., Hufnagle, D., Mulak, K., & Escudero, P. (2017). Speech Perception: Development. In *Reference Module in Neuroscience and Biobehavioral Psychology*. https://doi.org/10.1016/B978-0-12-809324-5.01904-0

De Houwer, A. (2017). Bilingual Language Acquisition. In *The Handbook of Child Language* (pp. 219–250). John Wiley & Sons, Ltd. https://doi.org/10.1111/b.9780631203124.1996.00009.x

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/article/10.3389/fpsyg.2014.00781

Dragojevic, M., Fasoli, F., Cramer, J., & Rakić, T. (2021). Toward a Century of Language Attitudes Research: Looking Back and Moving Forward. *Journal of Language and Social Psychology*, *40*(1), 60–79. https://doi.org/10.1177/0261927X20966714

Durrant, S., Delle Luche, C., Cattani, A., & Floccia, C. (2015). Monodialectal and multidialectal infants' representation of familiar words. *Journal of Child Language*, *42*(2), 447–465. https://doi.org/10.1017/S0305000914000063

Eriksson, M., Marschik, P., Tulviste, T., Almgren, M., Pérez-Pereira, M., Wehberg, S., Umek, L., Gayraud, F., Kovačević, M., & Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *The British Journal of Developmental Psychology*, *30*, 326–343. https://doi.org/10.1111/j.2044-835X.2011.02042.x

Floccia, C., Delle Luche, C., Durrant, S., Butler, J., & Goslin, J. (2012). Parent or community: Where do 20-month-olds exposed to two accents acquire their representation of words? *Cognition*, *124*(1), 95–100. https://doi.org/10.1016/j.cognition.2012.03.011

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. (2016). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *1*, 1–18. https://doi.org/10.1017/S0305000916000209

Garrison, H., Baudet, G., Breitfeld, E., Aberman, A., & Bergelson, E. (2020). Familiarity plays a small role in noun comprehension at 12–18 months. *Infancy, 25,* 458-477.https://doi.org/10.1111/infa.12333

Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-Five Years Using the Intermodal Preferential Looking Paradigm to Study Language Acquisition: What Have We Learned? *Perspectives on Psychological Science*, *8*(3), 26. https://doi.org/10.1177/1745691613484936

Johnsen, S. S. (2012). A diachronic account of phonological unnaturalness. *Phonology*, *29*(3), 505–531. https://doi.org/10.1017/S0952675712000243

Jusczyk, P. W., & Aslin, R. N. (1995). Infants′ Detection of the Sound Patterns of Words in Fluent Speech. *Cognitive Psychology*, *29*(1), 1–23. https://doi.org/10.1006/cogp.1995.1010

Kartushina, N. & Mayor, J. (2022). Coping with dialects from birth: Role of variability on infants' early language development. Insights from Norwegian dialects. *Developmental Science*.

Kartushina, N., Rosslund, A., & Mayor, J. (2021). Toddlers raised in multi-dialectal families learn words better in accented speech than those raised in monodialectal families. *Journal of Child Language*, 1-26. doi:10.1017/S0305000921000520

Mæhlum, B., & Røyneland, U. (2012). Det norske dialektlandskapet: Innføring i studiet av dialekter. In *NFFO*. Cappelen Damm akademisk. https://urn.nb.no/URN:NBN:no-nb_digibok_2018101677004

Mani, N., & Plunkett, K. (2010). Twelve-Month-Olds Know Their Cups From Their Keps and Tups. *Infancy*, *15*(5), 445–470. https://doi.org/10.1111/j.1532-7078.2009.00027.x

Mani, N., Schreiner, M. S., Brase, J., Köhler, K., Strassen, K., Postin, D., & Schultze, T. (2021). Sequential Bayes Factor designs in developmental research: Studies on early word learning. *Developmental Science*, *24*(4), e13097. https://doi.org/10.1111/desc.13097

ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed speech preference. *Advances in Methods and Practices in Psychological Science*, *3*, 24-52. https://doi.org/10.1177/2515245919900809

Mattock, K., Polka, L., Rvachew, S., & Krehm, M. (2010). The first steps in word learning are easier when the shoes fit: Comparing monolingual and bilingual infants. *Developmental Science*, *13*, 229–243. https://doi.org/10.1111/j.1467-7687.2009.00891.x

Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from CDI analysis. *Developmental Science*, *14*(4), 769-785.

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2021). *BayesFactor: Computation of Bayes Factors for Common Designs.* R package version 0.9.12-4.3. Retrieved from https://CRAN.R-project.org/package=BayesFactor.

Munoz, L., Mayor, J., & Kartushina, N. (in prep). Dialect exposure and vocabulary development in Norwegian infants: Insights from parental reports.

Norwegian Labour and Welfare Administration. (2021, February 5). *Foreldrepenger, engangsstønad og svangerskapspenger*. NAV. Retrieved October 2021, from https://www.nav.no/no/nav-og-samfunn/statistikk/familie-statistikk/foreldrepenger-engangsstonad-og-svangerskapspenger.

R Core Team. (2012). *R: A language and environment for statistical computing*. Retrieved from: http://www.R-project.org.html

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. Developmental science, 12(2), 339-349. https://doi.org/10.1111/j.1467-7687.2008.00786.x

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. Infancy, 15(6), 608-635. https://doi.org/10.1111/j.1532-7078.2010.00033.x

Schmale, R., Hollich, G., & Seidl, A. (2011). Contending with foreign accent in early word learning. *Journal of Child Language*, *38*(5), 1096–1108. https://doi.org/10.1017/S0305000910000619

Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic bulletin & review, 25(1)*, 128-142. https://doi.org/10.3758/s13423-017-1230-y

Schönbrodt, F. D., Wagenmakers, E. J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological methods*, *22*(2), 322–339. https://doi.org/10.1037/met0000061

Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., & Jørgensen, R. N. (2014). The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language*, *34*(1), 3–23. https://doi.org/10.1177/0142723713510997

Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, *51*(2), 173–189. https://doi.org/10.1016/j.jml.2004.04.004

Swingley, D. (2003). Phonetic Detail in the Developing Lexicon. *Language and Speech*, *46*(2–3), 265–294. https://doi.org/10.1177/00238309030460021001

van der Feest, S. V. H., & Johnson, E. K. (2016). Input-driven differences in toddlers’ perception of a disappearing phonological contrast. *Language Acquisition*, *23*(2), 89–111. https://doi.org/10.1080/10489223.2015.1047096 van Heugten, M., & Johnson, E. K. (2014). Learning to contend with accents in infancy: Benefits of brief speaker exposure. *Journal of Experimental Psychology: General*, *143*(1), 340. https://doi.org/10.1037/a0032192

van Heugten, M., & Johnson, E. K. (2017). Input matters: Multi-accent language exposure affects word form recognition in infancy. *The Journal of the Acoustical Society of America*, *142*(2), EL196–EL200. https://doi.org/10.1121/1.4997604

Venås, K., & Skjekkeland, M. (2021). Dialekter i Norge. In *Store norske leksikon*. http://snl.no/dialekter_i_Norge

Von Holzen, K., & Bergmann, C. (2021). The development of infants' responses to mispronunciations: A meta-analysis. *Developmental Psychology*, *57*(1), 1–18. https://doi.org/10.1037/dev0001141

Wagenmakers, E.-J., Etz, A., Gronau, Q., & Dablander, F. (2018, December 20). The Single Most Prevalent Misinterpretation of Bayes' Rule. *Bayesian Spectacles*. https://www.bayesianspectacles.org/the-single-most-prevalent-misinterpretation-of-bayes-rule/

Wagenmakers, E.-J., Morey, R.D., & Lee, M.D. (2016). Bayesian benefits for the pragmatic researcher. Current Directions in Psychological Science, 25(3), 169–176. https://doi.org/10.1177/0963721416643289

Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism in infancy: First steps in perception and comprehension. *Trends in Cognitive Sciences*, *12*(4), 144–151. https://doi.org/10.1016/j.tics.2008.01.008

White, K. S., & Aslin, R. N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*, *14*(2), 372–384. https://doi.org/10.1111/j.1467-7687.2010.00986.x

Zettersten, M., Bergey, C. A., Bhatt, N., Boyce, V., Braginsky, M., Carstensen, A., deMayo, B., Kachergis, G., Lewis, M., Long, B., MacDonald, K., Mankewitz, J., Meylan, S. C., Saleh, A. N., Schneider, R. M., Tsui, A., Uner, S., Xu, T. L., Yurovsky, D., & Frank, M.C. (2021). Peekbank: Exploring children's word recognition through an open, large-scale repository for developmental eye-tracking data. Proceedings of the 43rd Annual Conference of the Cognitive Science Society. https://psyarxiv.com/ep693