Assessing Early Word Comprehension

A comparison of Parental Reports and Infants' Performance in a Pupillometry Study.

Marisol Hernández Ortega

Submitted as a master thesis at the Department of Psychology, University of Oslo

Spring 2022

Declaration

This thesis is submitted to the University of Oslo in completion of the requirements for a master thesis in psychology. The data obtained to perform the presented analyses stems from the research project called: The Developmental Trajectory of Early Word Representation in 8- 12-Month-Old Infants: Insights from Pupillometry, conducted by Julien Mayor and Natalia Kartushina. A Project on which I contributed from assisting in the selection and design of the stimuli to the recruitment and coordination of the families who participated in the experiment.

I, Marisol Hernandez Ortega, declare this work to be of my own authorship except when acknowledged.

Acknowledgments

First and foremost, I would like to thank my supervisors Natalia Kartushina and Julien Mayor, for sharing their valuable knowledge through my years as a master student and for their essential contributions to this thesis.

I would also like to thank Ph.D. candidate and former classmate Audun Rosslund, for his willing to contribute to my work with discussions and ideas, always being friendly and accessible during this research process.

I wish to extend my most sincere gratitude to my children Lía and Elliot; and to my husband Alexander, for their unconditional love and patience during this exhaustive period in our lives. My love for you will always be the biggest motivation for whatever the next challenge shall be.

Special thanks to family and friends who always have been cheerful and supportive in my academical achievements and my life abroad, I appreciate every single one of you. Finally, thanks to all the families and participants that gave this study a thought!

Sincerely,

Marisol Hernández Ortega.

Abstract

Author: Marisol Hernandez Ortega

Title: Assessing Early Word Comprehension: A Comparison of Parental Reports and Infants' Performance in a Pupillometry Study

Supervisors: Natalia Kartushina and Julien Mayor

Background: Early word comprehension has been extensively studied in the psychological and linguistic domains to gain insights in the processes involved in language development. One of the most used tools to measure word comprehension in very young children are parental reports. However, the use of direct measures as a convergent method to screen the accuracy of these reports is scarce, for that reason, the present study explores the effectiveness of parental reports of their infants' early word knowledge in comparison with the outcomes of pupil dilation analyses as a direct measure of word comprehension

Method: In this study we analyzed parental reports of their infants' early word comprehension by comparing their responses from the MacArthur-Bates Communicative Inventories questionnaire in its Words and Gestures version, to their infants' measures of pupil dilation, as indexes of word comprehension. For our statistical analyses we opted for mixed-effect models and performed linear mixed model analysis fit by maximum likelihood, using R program for statistical computing.

Results: Our study do not reveal robust evidence of parents being good assessors of their infants' word comprehension.

Conclusions: Our findings reinforce the suggestions of using supplemental measures in the assessment of infants' word knowledge, rather than the use of parental reports as a sole measure. However, Norwegian parents might be good assessors of their children's word comprehension, yet more research in older children and with more familiar items are suggested to consolidate parental reports as a complementary and/or convergent measure of infants' word comprehension.

Key words: Pupillometry, Eye-tracker, developmental psychology, parental reports, early word comprehension.

Table of contents

List of abbreviations

AIC = Akaike Information Criterion

BIC = Bayesian Information Criterion

CDI = MacArthur-Bates Communicative Development Inventories

CDI-WG = MacArthur-Bates Communicative Development Inventories – Words and Gestures

CDI-WS = MacArthur-Bates Communicative Development Inventories – Words and Sentences

CDI-III = MacArthur-Bates Communicative Development Inventories Toddler Form

dB = Decibel

df = Degrees of freedom

df Den. = Degrees of freedom - Denominator

df Num. = Number of Degrees of Freedom

E.g. = Example given

ERP = Event-Related Brain Potentials

Et al. = et alia – “and others”

F = Value of distribution

HPP = Head-turn Preference Procedure

Hz. = Hertz

Id. = Identification

I.e. = In example

IPL = Intermodal Preferential Looking

LWL = Looking While Listening Task

M = Mean

Max. = Maximum

Med. = Median

Min. = Minimum

Ms = Milliseconds

NSD = Norwegian Centre for Research Data

p = Probability value

Q1 = Quartile 1

Q3 = Quartile 3

S = Second

SD = Standard Deviation

SE = Standard Error

Sq = Square

t = Value of the difference between population means

Trasc. = Transcription

Studies of early word comprehension

Studies of early lexical development are central in educational and psychological research, first, because word comprehension is associated with subsequent vocabulary growth (Bleses et al, 2016; Friend et al, 2012), which is essential for a child's later educational accomplishments and favorable academic outcomes (Bleses et al., 2016; Morgan et al., 2015; Simonsen et al., 2014); and second, for its correlation with neurobiological aspects (Kuhl, 2004) and cognitive development (Fitch, 2011).

To date, word comprehension in young children has been the target of numerous empirical investigations for a better understanding of the constructs that concern language development, with the purpose to provide effective assessment and give the opportunity to detect early language delays or other difficulties related to the children's development of speech and other linguistic skills. However, there is limited scientific data about the assessment of early word representations on infants under the age of 12 months (Bergelson & Aslin, 2017; Syrnyk & Meints, 2017), and even less in Norwegian children (Simonsen et al., 2014; Kartushina & Mayor, 2019). Considerable research should thus be dedicated to gaining insights into the process and mechanisms underlying early vocabulary knowledge in Norwegian infants.

Early Stages of Language Acquisition

The human cognitive and perceptual abilities underlying language development are sensitive to speech features even at a prenatal stage (Ghio et al., 2021). Soon after birth, infants already show language processing skills (e.g., 0- 6-days old newborns recognizing sounds and speech variations from their mother tongue (Gervain, 2015; Yeung & Werker, 2009). Furthermore, at ages between 3 to 6 months infants have already developed sophisticated learning qualities, so the first milestones of word comprehension occur; first, infants start recognizing sound patterns (Jusczyk, 1997; Tincoff & Jusczyk, 1999); and second, they associate these aspects to objects or events (Haith, 1994; Swingley, 2009). However, these milestones are just building blocks to other processes that take place during the first stages of language acquisition. Whereas word comprehension and vocabulary growth might at first glance seem like a simple *word – meaning* learning task, they are processes conjoined to other social and neurological aspects that make language acquisition far from straightforward.

In a review of the literature, Gervain and Mehler (2010) dispute the centralization of language acquisition as a merely biological prospect. In turn, they propose an integrative approach, merging linguistics, biological and social factors (among other cognitive considerations e.g., attention and memory) that, in line with other studies, indicate that early lexical comprehension relies on infants' complementary abilities. For example, the use of contextual referents (Arias-Trejo & Plunkett, 2010), semantic structures (Bergelson & Swingley, 2012), phonological features (Tamasi et al., 2019), and frequency cues (Kartushina & Mayor, 2019).

To have a better understanding of the different features that help infants categorize information from lexical input, researchers have aimed to study the earliest stages of word comprehension on infants, which to date has been documented as early as at the age of 4.5 months (e.g., recognizing their own names in Mandel et al., 1995). Furthermore, in recent investigations it has been identified that infants at ages between 6- 9-months understand some words for food (e.g., "banana" in Bergelson & Swingley, 2013), body parts (e.g., "hair" in Bergelson & Swingley, 2012) and common verbs (e.g., "hear" and "see" in Bergelson & Swingley, 2012; Bergelson & Aslin, 2017). However, studies of word comprehension on infants before their first birthday have been faced with the challenge of measuring a process that is not readily observable. Whereas articulated utterances can be quantifiable by monitoring a child's word production, the assessment of comprehension, i.e., what infants grasp from spoken language before they produce their first words, is a complex task, given infants' lack of verbal feedback. Thus, research in this domain has approached the parents' information about their infants' word understanding (Houston-Price et al., 2007; Styles & Plunkett, 2008; Syrnyk & Meints, 2017). Nonetheless, concerns about this method have been rising. Can parents be good evaluators of their children's early word knowledge?

Previous Documentation of Parental Reports Assessing Children's Early Word Comprehension

Historically, the assessment of word understanding at early ages has emphasized the children's first milestones during their developmental journey. For example, scientists first addressed infants' spurts in their comprehension of receptive language through longitudinal observations, documenting the way they responded to directed words in their own environments, with records of gestures and looking behaviors (Bloom, 1970; Camaioni et al., 1991; Miller, 1972). Yet, the use of these techniques over time have been described as to be highly time consuming (Bauer et al., 2002; Reznick & Goldsmith, 1989) hence, not suitable for large scale applications. To date, however, studies of word comprehension on infants under one year rely on parental information about their child's vocabulary knowledge, mainly because parental reports provide an accessible methodological approach to almost any kind of pediatric examinations, including the assessment of speech development in clinical settings and for academic research purposes (Sachse & Von Suchodoletz, 2008; Yoder et al., 1997). Parental reports offer information collected over an extended period of time and across diverse contexts (Bauer et al., 2002; Feldman et al., 2000; Fenson et al., 1994; Fenson et al., 2000), providing researchers and clinicians an approach to the child's word comprehension and word production.

One of the most frequently used parental report tools to measure language development in very young children is the MacArthur-Bates Communicative Development Inventories, also known as CDIs. CDIs are a tool used to assess vocabulary comprehension and production in children between 8- 37-months of age (Fenson et al., 1994; Fenson et al., 2000), where the parents or caregivers can fill in a prespecified list of words or/and sentences their young children can already understand and/or produce (Mayor & Mani, 2018). This instrument can be found in their three existing versions: The CDI Words and Gestures (CDI-WG) designed for infants between the ages of 8- 16-months, assesses early vocabulary comprehension and production including communicative gestures; The CDI Words and Sentences (CDI-WS) designed for young children in ages from 16- 30-months, assesses productive vocabulary and early grammar; and The CDI-III (Fenson et al., 2000) assesses language development in toddlers from 30- 37-months of age, presented in short forms.

The CDI's, in their multiple versions in diverse languages (e.g., in Norwegian: Simonsen et al., 2014) offers the practical advantages of being low-cost, can be filled out in a home environment, does not require any clinical training and is independent from children's cooperation. However, when it comes to the assessment of word comprehension of children under the age of 12-months, the reliability and precision of parents' intuitions when filling out the CDI's have been considered as not very objective. For instance, Tomasello and Mervis (1994) argued that parents can easily tend to over-interpret their infants' word comprehension, biased by what could just be the child recognizing familiar sounds or gestures and thus responding to non-verbal information that indicates the word referent. Additionally, other arguments affirm that diverse predispositions of the infant's family, such as ethnicity, literacy obstacles, or social class position, affect the accuracy of how parents perceive their infants' word comprehension (Hart & Risley, 1995; Feldman et al., 2000) and hence, provide imprecise reports.

Socioeconomic Status and its Relationship with Children Vocabulary Scores

The families' socioeconomic status (SES) refers to the resources that represent their social position according to their financial capital and income (Bradley & Corwyn, 2002). Additionally, other indicators aside from wealth and social status such as parental education level and occupation have also been found as mediating factors in language development at early ages; these arguments have encouraged researchers to consider familial SES in studies of children's cognitive abilities including language proficiency (Rindermann & Baumeister, 2015). In line with these statements, recent studies have pointed at SES indexes and its variations between families, discussing that SES have a tendency to influence the scores of children's vocabulary development measures, where the families with lower SES rates might be at disadvantage and more likely to be "at risk" for lower vocabulary development, than families with higher SES (Arriaga et al., 1998; Fernald et al., 2013; Hart & Risley, 1995; Hoff, 2003). However, although the scores on production checklists tend to be lower for children coming from families with lower rates of SES, their parents don't necessarily under-report comprehension. As an example, Feldman and colleagues (2000) studied parental reports of vocabulary comprehension in children between 8-16-months of age, in a large socio-demographically diverse sample (N = 2,156) and their results showed that caregivers with low levels of income seemed to rather overestimate their child's comprehension abilities in comparison to caregivers with higher levels, to which authors discussed could be the result of potential biases related to the parents' education level, for example the misunderstanding of the instructions or the fact that infants' word comprehension demands the interpretation of what the infant understands of spoken language instead of taking account the contextual references that lead to over interpretation.

Frequency Data vs. Word Familiarity

Earlier analyses have questioned the standards from the CDI-WG regarding the general expectations of words children should be exposed to at ages from 8- 16-months, versus the infants' familiarity with these word referents. Previous research in older children (e.g., 19- 24-months in Fernald & Weisleder, 2013) show that the continuity by which an object is presented simultaneously with its word referent to an infant, shapes word learning, suggesting that the family's linguistic input is crucial for the child's speech development. However, some analyses discuss that infants at ages under 12-months can constrain early word meaning without continuously being exposed to the word referent, as discussed in Bergelson & Swingley (2012)'s study, where authors examined word recognition in English speaking infants between 6- 9-months of age. In this study, authors included parental information of the words infants' were familiar with. Their outcomes showed that parental reports of words used frequently in the child's presence did not correlate significantly with the words that were shown as understood by the infant. Furthermore, researchers argue that at such early ages, infants' word recognition is related to the meaning of the words, sustaining that semantics learnt by experience have a higher impact on children's word learning rather than the regularity of occurrence where items are presented simultaneously with their label.

In contrast, Syrnyk & Meints (2017) tested word knowledge in English speaking 9-month-olds in three different conditions: controlling for (a) words that were reported as understood by the parents. In this condition, researchers used a wider range if words, including 8 additional items in a parent tailored fashion. These items were words involved in specific infant-related actions and common routines (e.g., going to sleep: "*night-night*"); (b) words that were expected to be understood by age-related frequency data; and (c) words reported from the parents as "*unknown"* for the children. In their results, authors discuss that at ages around nine months, parents may be better assessors of their own sons or daughters' word knowledge, than what the general expectations from what word frequency data standards would predict, considering that infants showed little- to no-comprehension of the words that might be expected to be understood from the CDI-WG data (condition b). Infants did not display word comprehension for words reported from the parents as not known by the infants (condition c) aligning with their statement of parents as good indicators of their infants' word recognition. To conclude their study, authors suggest that parental assessment of infants' word knowledge is effective when using tailored, individualized lists.

Direct Measures of Word Comprehension.

Direct measures of word understanding are performed by directly measuring the infants' responses to language stimulus in time locked trials (Sachse & von Suchodoletz, 2008). To date the most used approaches are the looking-based measures, as they record infants' looking behaviors or gaze fixations under observable, time-coursed settings while processing visual stimuli. E.g., the head-turn preference procedure (HPP) (Nelson et al., 1995) which consists of tracking infants' head-turns as a response of interest to a directed stimuli, and have been conducted to investigate word comprehension in 6- 7.5-month-olds in Jusczyk and Aslin (1995); 8-month-olds in Jusczyk and Hohne (1997); 7.5- and 10.5- month-olds in Singh et al. (2004); 8-month-olds in McMurray and Aslin (2005); and 11-month-olds in Swingley (2005).

Other examples are the intermodal preferential looking paradigm (IPL) (Golinkoff et al., 1987) that explore infants' gazes between two images presented one at a time (e.g., a target and a disctracter), revealing comprehension to the item (word referent) obtaining larger fixations, and its variant: the Looking-while-listening task (LWL) (Fernald et al., 2008; Swingley, 2011), where two or more pictures of familiar objects are presented simultaneously, hearing the auditory label of one of the depicted objects; have been used to study lexical representations of 6-month-old babies in Tincoff & Jusczyk (1999; 2012); 8- 9-month-olds in Schafer (2005); 6- 9-month-olds in Bergelson and Swingley (2012); 6- 9-month-olds in Bergelson and Swingley (2013); 6- 14-month-olds in Bergelson and Aslin (2017); 9-month-olds in Syrnyk and Meints (2017) and 6- 9-month-olds in Kartushina and Mayor (2019).

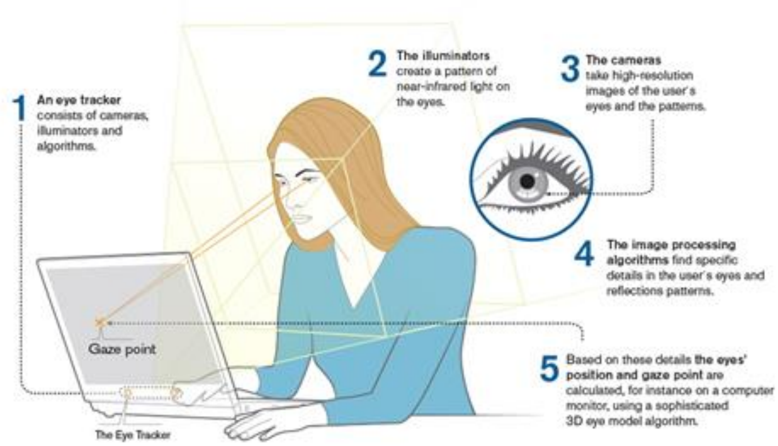
Other methods, such as event-related brain potentials (ERPs), which measure brain responses to different types of stimuli, have been used to investigate word-object representations (Friedrich & Friederici, 2011; Parise & Csibra, 2012), providing reliable evidence in the assessment of early word recognition in infants. In summary, the use of novel techniques with the help of technology and its increasing sophistication, have been getting more precise in terms of measuring children's processes and features involved in word learning during their first months of life.

Measures of Pupil Dilation

The recent popularity of the use of eye-trackers in the study of language acquisition has given scholars the opportunity to approach the newly re-discovered research method based on measures of the human pupil dilation. Pupil dilation analysis consist of collecting and measuring changes in pupil size in response to stimulus, over a period of time (Hepach & Westermann, 2016). In this paradigm, pupil diameter is tightly related to cognitive load and mental effort (Beatty, 1982; Eckstein et al., 2017;2019; Hess & Polt, 1964; Schmidtke, 2018; van der Wel & van Steenbergen, 2018).

Figure 1

Functional Description of an On-Screen Eye-Tracker Device. Image Retrieved from the Tobii Pro Website, Shared with Permission.



Pupil dilation in humans

The pupillary responses in the human eye are involuntary changes in the pupil diameter, compound by the sphincter pupillae muscle in the iris to evoke pupil constriction, and the dilator pupillae muscle of the iris to provide pupil dilation (McDougal & Gamlin, 2015). This process is determined by diverse circumstances such as luminance, age, corneal flux-density (Napieralski & Rynkiewicz, 2019) or other influences of light passing through the eye's biological structure (Eckstein et al., 2017; Zele & Gamlin, 2020). However, changes occurring in pupil sizes are also modulated by the brain's locus coeruleus-norepinephrine system, which commands the regulation of physical arousal and is central in functions that underlie behavior and cognitive functions (Sara, 2009). Hence, the pupils respond to cognitive demands and mental activity and not exclusively photo-sensory stimuli, as explained in Figure 2.

Figure 2

Illustration of the Pupillary Light Reflex. The Balance Between the Dilator (Left) and Sphincter (Right) Pupillae Muscle Contraction Dictate Pupil Size. Adapted from Eckstein and Colleagues (2017)

When it comes to young children, recent investigations have shown that infants display greater degrees of pupil dilation while performing tasks that demand high cognitive effort (Hepach & Westermann, 2016; Hochmann, 2013; Tamási et al., 2017), for example, by experiencing unexpected events (8-month-olds in Jackson & Sirois, 2009; 10-month-olds in Sirois & Jackson, 2012), arousal and distress (6- 12-month-olds in Geangu et al., 2011), and under unfamiliar circumstances (6- and 12-month-olds in Gredebäck & Melinder, 2010). Furthermore, and more relevant to our study, Tamási and colleagues (2019) described pupil dilation measures as a promising supportive method to existing research paradigms to explore early lexical representations, due to the positive correlations between the results of pupil dilation and speech development in their studies (Tamási et al., 2017; 2019), revealing that pupillary-response measures are sensitive to phonological features (e.g., phonological mismatch) in line with other research findings in younger children, indicating sensitivity to acoustic differences in 3- and 6- month-olds (Hochmann & Papeo, 2014); and segmentation abilities in continuous speech in 5- month-olds (Fló, 2021). Such findings thus suggest that measures of pupil dilation offer the sensitivity to detect known or unknown words in infants.

Mechanisms Underlying Experimental Research of Pupil Dilation

Pupil data for cognitive analyses can be collected in inexpensive and non-invasive ways. However, the methodological challenges underlying experimental designs with measures of pupil dilation should be carefully prepared (Hayes & Petrov, 2016). For instance, the luminance in the research room and research screen must be in constant conditions to prevent changes in pupil sizes induced by light variations. Considering this, no natural light should reach the experiment room, and the mean luminance in all images depicted in the testing screen may not differ (Hepach & Westermann, 2016; Schmidtke, 2017). Furthermore, gaps in pupil measurements (e.g., the infant blinks or looks away) leads to data loss and fluctuations in pupil sizes. Such fluctuations are a source of noise that interferes with the measures of interest (Mathôt et al., 2018), for that reason it is considered that pupil dilation analyses work best when the infants' fixations to the screen are stable, that way researchers obtain a pre-trial baseline period to start measuring pupil size changes (Beatty & Lucero-Wagoner, 2000). Thus, it has been estimated that a mean value of 100 ms pre-trial interval is adequate to obtain pupil size average measures for analyses of pupil dilation, in research with infants under one year (Geangu et al., 2011; Zhang et al., 2019).

Baseline correction

Pupil size changes determined by the onset of an experimental manipulation are statistically stronger when they are baseline corrected, which is in other words, the comparison of changes in pupil sizes after a stimulus onset, to its size in a prior established baseline. The most recommended way to apply baseline correction, is by subtracting the baseline value to the pupil score post stimuli (e.g., corrected pupil size = post-stimuli pupil size – baseline, in Mathôt et al., 2018). It is also suggested to establish a baseline in between trials, to avoid fluctuating data during an experiment, since baseline diameters change due attention lapses or fatigue (Schmidtke, 2017).

The Present Study

This thesis examines parental reports and their efficiency to estimate their infants' early word comprehension: we investigated whether Norwegian parents are good assessors of their child's word comprehension in ages between 8- 12-months by measuring parental reports of their child's word comprehension and the infants' performance of word comprehension in a pupillometry study.

Word comprehension indexes

We analyzed infants' word understanding with the use of an eye-tracker for pupil dilation analyses, measuring the infants' pupil size in the following conditions: when the picture of a target object is named with a correct label (e.g., apple - apple) versus when the picture of a target object is named with a non-matching and contextually unrelated label (e.g., apple - pants). We expected larger pupil sizes in unrelated as compared to matching trials, indicating infants' word comprehension of the matching label, based on the construct from previous studies where a greater degree of pupil dilation in infants is linked to increased cognitive loads (Tamási et al., 2017; van der Wel & van Steenbergen, 2018). Hence, lack of word comprehension would display similar pupil sizes on both conditions as the infant do not actively try to relate a known word object with an unmatching label.

Parental Reports

The parental reports were collected via the Norwegian adaptation of the CDI-WG questionnaire (Simonsen et al., 2014), a checklist consisting of 393 words vocabulary items (including our 16 target objects) filled out by the parents of the children that matched our eligible criteria (for details see the participants' section). The questionnaires were submitted to each participant with the specification that only one parent per child must answer. Yet, relevant information about both parents (including gender, highest level of education and estimated percentage of time spent with the infant) was required. The responding parent could choose between the options *understands* and *understands and produces*, for each word in the checklist. In case their children did not show evidence for any of these options, the instruction was to leave the question unanswered.

To our analyses, we explored the word items reported from the responding parent as *understands* and *understands and produces* for the child and compared them with the words that were left unanswered (not understood), analyzing if the words reported as understood have higher pupil-size differences as compared to the items that were reported as non-understood.

Specific Aims

The current study has two specific aims. One, to make a contribution to the scarce research of word comprehension in Norwegian infants at its earliest stages (8- 12- months-old in Kartushina & Mayor, 2019). And two, to examine the relationship between parental reports and direct measures of word comprehension in early childhood. Previous research by Syrnyk & Meints (2017), in an IPL study showed alignment between word comprehension in 9-month-old English-speaking infants and words reported as understood from their parents. In this study, we tried to replicate this measure, yet this time with the outcomes of pupil dilation analyses, similar to Tamási et al (2017; 2019).

Hypothesis

We hypothesize that parents will be reliable informers of their child's lexical knowledge as described in Fenson et al. (1994), Styles & Plunkett (2008) and Syrnyk & Meints (2017). We thus predict a significant correlation between words reported as understood vs non-understood and the children's word recognition indicated by changes in the pupil dilation size.

Method

Participants

Parents of 188 Norwegian children from the Oslo region in Norway were contacted to participate in the study, selected from the list of the Norwegian National Population Registry (Folkeregisteret). The invitations were distributed via postal mail with the instruction of signing up and answering our registering survey. After signing up and having consented to their participation, parents of children who matched our eligible criteria were asked to fill out a language background questionnaire and were recruited to participate. Once our appointment to meet up at the laboratory was set up, the CDI-WG (Simonsen et al., 2014) was sent to the parents via e-mail with a link to the *nettskjema* platform from the University of Oslo. They were asked to fill out the questionnaire in a period not longer than one-week prior to the experiment day.

The eligibility criteria of the children were the following: (i) birth at full-term (gestational weeks greater than 37) with an age between 8- 12-months; (ii) minimum 90% exposure to Norwegian language; (iii) no developmental delays or history of visual/hearing impairments.

The total sample of recruited infants who matched our eligible criteria was *n* = 64, of which 39 were boys and 25 were girls, with a mean age of 339.6 days old (*SD* = 20). Parents' SES was indexed by the mother's highest education level (as in Feldman et al., 2000), ranging from 0 (*primary school*) to 5 (*doctoral degree*) with a mean score of 3.73 (*SD* = .60). Data from other 26 infants was collected but not included in the final sample because they did not match our analyses' eligibility criteria (see data processing section for details).

The experiment took place at the BabyLing eye-tracking laboratory, located at the facilities of the faculty of psychology in the University of Oslo.

Ethics

The present study was approved by the ethics committee at the department of Psychology and by the Norwegian Centre for Research Data (NSD) Ref. 56312.

Design

The experiment is a within-participant design, where eight highly familiar words have been presented to a child with the picture of the referent object (e.g., an apple) followed by an audio message prompting the child to look at it (e.g., "look at the " /object/). Each trial has been displayed in 3 conditions, with: (a) the matching label (e.g., apple – apple), (b) a non-matching but contextually related label (e.g., apple – table), (c) a non-matching and contextually unrelated label (e.g., apple – pants). Making a total of 24 trials. The analyses and results from condition (b) are beyond the scope of the current project and will be presented in a different work. Making a total of 16 trials for the present study.

We examined the infants' pupil size with an eye-tracker on the conditions (a) and (c): when the label is matching an image associated with a matching word, versus when the label is mismatching the image. The measures of pupil size obtained by the eye-tracker were corrected for inter-trial and inter-subject variation (e.g., per-item and for each participant). We corrected by subtracting pupil size at pre-naming period (e.g., a mean value of a 100 ms interval before the onset of the auditory label) from the pupil size at post-naming period, this way, we obtained average baseline-corrected data from unrelated and matching trials. Then, we computed the difference between average baseline-corrected pupil sizes at unrelated and matching trials to obtain our pupil size measures of interest (e.g., pupil size difference = unrelated - matching).

The collected data from our dependent variable (pupil size measures) were compared to the words reported as understood and non-understood by the parents (in our language comprehension questionnaire), to examine whether items reported as understood by the parents have higher difference in pupil sizes as compared to the items that were reported as not understood for the infant.

Randomization

All trials were pseudo-randomized, the same object could be presented only after that two or more trials had been passed from the last time it was presented.

Stimuli

Eight lexical items were extracted for our study from the Norwegian version of the CDI-WG (See table 1). Our eight target words were selected following the criteria of (i) having the highest rate of familiarity from the basis of the CDI normative data for 9-11- month- old Norwegian babies, using the WordBank word database of children's vocabulary development (Frank et al., 2017); and (ii) by having the highest rates of word frequency from CHILDES talk-bank system (MacWhinney, 2019), as shown in Table 1. Four of our target words were food/kitchen-related objects, while the other four were body/cleaning-routine related. Our other eight lexical items were assessed under the condition of being unrelated words (must not be semantically related nor contextually related), matched only by frequency and divided in body/cleaning-routine and food/kitchen-related objects (four for each of them).

Table 1

List of Matching and Unrelated Labels

Matching words								
English	banana	bread	apple	cookie	foot	belly	hair	leg
Norwegian	banan	brød	eple	kjeks	fot	mage	hår	bein
Phonetic transc.	bɑnɑ:n	brø:	eplə	çeks	fu:t	mɑ:gə	ho:r	bæjn

Frequency	107	26	40	20	11	6	47	16
CDI	46	32	19	17	23	19	17	7
Unrelated words								
English	doll	telephone	pants	picture	pacifier	stroller	light	box
Norwegian	dukke	telefon	bukse	bilde	smokk	barnevogn	lys	eske
Phonetic transc.	du`k:ə	teləfo:´n	buksə	bi´ldə	smok:	ba:`rnəvåŋn	ly:s	e`skə
Frequency	36	20	18	57	6	28	13	2
CDI	20	27	14	13	43	29	29	8

Visual stimuli

Pictures of the eight highly familiar target word objects (e.g., banana, bread, apple, cookie, leg, foot, hair, belly), were displayed one at a time, always preceded by a grayscale isoluminant slide before presenting the target stimuli picture, all of them with an identical luminance value throughout the trials. The labels of the matching and unrelated conditions were presented randomly as auditory stimuli to each of the target object images (For details see description section). The pictures were depicted in a light-gray background with a 1280 x 1024 resolution (1280 on the x axis, 38 bits, 60Hz), with a constant luminance in all stimuli to avoid changes in pupil-size (Eckstein et al., 2017; Zhang et al., 2019).

Figure 3

Example of Visual Stimuli for Target Object "Apple"

Audio stimuli

The audio stimuli were recorded by a female native speaker of Easter Norwegian dialect was recorded while reading at a slow speed and child-directed fashion 26 sentences, 24 of which included the eight lexical target words used for the trial, conjoined with four types of carrier sentences: E.g., “Look! A /object/!”, “look at the /object/”, “oh! A /object/!” and “here is the /object/”. (The two other sentences used in the task were one *welcome* trial to introduce the infant to the task and one *thank you* trial to conclude the experiment). The carrier sentences were used to prompt the infant to look after each word referent during the trial. We processed the records using the Praat computer program (Boersma & Weenink, 2020) to remove additional noise and equalize the intensity of the target words with the prompting sentences. The parameters for our recordings were the following: 16 bits, 2 two channels, 44kHz. The stimuli average amplitude was equalized and set to 70dB as in previous research (Bergelson & Swingley, 2012; Kartushina & Mayor, 2019).

Description

The participants were received by a Norwegian native speaker at the BabyLing laboratory, then the parent(s) received a brief explanation of the task and then sign a consent form, to afterwards be followed by the experimenter to the eye-tracking room. The infant sat on one of the parent's lap facing the experimental computer's screen. The parent was required to wear headphones and listen to masking music to avoid potential interfering biases during the task and was asked not to talk to the baby, point to the screen or shift position. The experimenter was placed behind the participants in a position where the participant could not get distracted by her presence, yet she was able to monitor the infants' gaze behavior through a control screen. In this experiment, the objects were presented on three conditions, each containing eight trials (24 trials in total). First, with a matching label; second, a non-matching but contextually related label (which as previously mentioned, will not be presented in this study); and third, a non-matching and non-contextually unrelated label.

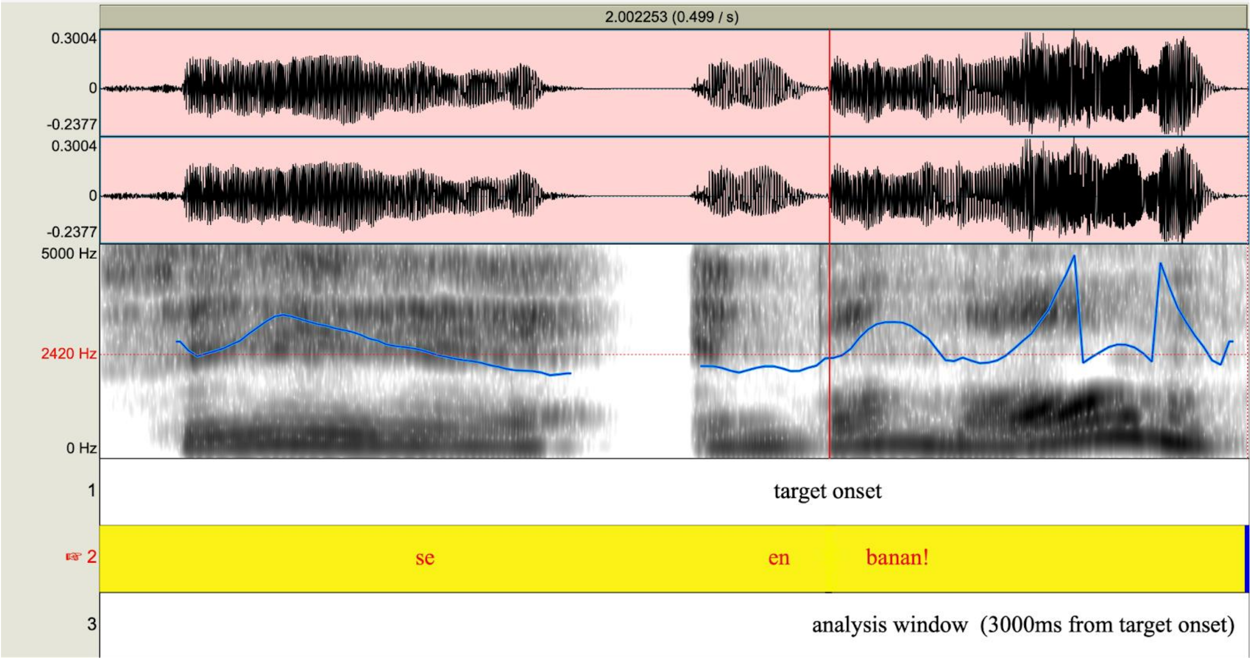
Gaze position and pupil size changes were monitored by an Eye-link 1000 eye-tracker, 500Hz sampling rate (monocular), at infant rate calibration and a 1280 x 1024 pixels screen resolution. The images were displayed in a screen with luminosity matching the pictures of the target objects presented in the stimuli. The experiment with an automated 5-point calibration procedure (slow version), which was followed by a gray isoluminant slide in a time-baseline of 1500 milliseconds. The visual stimuli were presented one image at a time, lasting 1000 ms in silence before the auditory label was played (with an average amplitude of 65 dB through two speakers located at the left and right sides of the screen). The last 100 ms before the auditory word onset was played were used as a baseline window, whereas our window of analyses for pupil dilation analyses was of 3000 ms from auditory target word onset. In case the infant did not fixate her gaze on the gray slide prior to each trial for 200 ms minimum, an attention getter (e.g., a spinning wheel on a gray background with the sound of birds tweeting) would come up to bring back the child's gaze to the screen. Once the infant looked back at the screen for more than 200 ms, the image of a target object was presented as previously explained and remained in the screen for 3000 ms after the target word onset.

Note that the audio sentences vary in length according to the nature of the word and carrier sentences, the mean pupil dilation in post-naming windows will be calculated on a trial-wise basis, where each trial will serve as its own baseline (see Figure 4 for an example).

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

Figure 4

Illustration of the Timeline for the Auditory Trial "Banana". The Sentence is "Look! A Banana!" - "Se! En Banan!"



Questionnaire

Parents rated their children's comprehension skills by completing the Norwegian adaptation of the CDI-WG. The vocabulary size of the participants was computed by registering the overall comprehension scores per child as shown in Table 3.

Table 3

Vocabulary Comprehension Scores Retrieved from the Parents' Responses of the Norwegian CDI-WG.

	Min	Med	Max	*M*	*SD*
Understood words	4	48	205	56	41

Additionally, the parents had the option of rating the familiarity of the infant with the word referents used as target items in our experiment, by indicating how often these eight words have been used in the child's presence, similar to previous research (Bergelson & Swingley, 2015; Houston-Price et al., 2007; Syrnyk & Meints; 2017) with the purpose to explore if word familiarity has an effect in the parents' predictions of their infants' word comprehension. These words were rated from zero (*never used in the child's presence*), to five (*extremely often used in the child's presence*). In our study, the most used word was "bread" with a mean score of 3.17 (*SD* = 1.06) and the least used word was "cookie" with a mean score of .77 (*SD* = .81). The percentages and frequency of the words reported as never being mentioned in the infants' presence are described in Table 4.

Table 4

Target Words Reported as Non-Heard for the Children

	Banana	Bread	Apple	Cookie	Leg	Foot	Hair	Belly
Nr. of participants	1	1	4	28	7	2	14	1
%	1.6	1.6	6.3	43.8	10.9	3.1	21.9	1.6

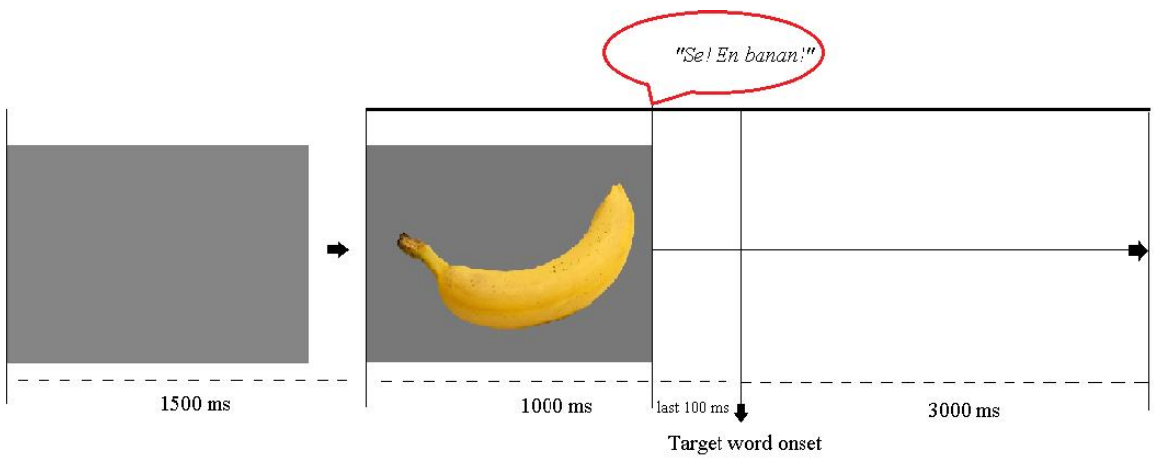
Data Processing

The final sample of participants was retained by applying the following exclusion criteria to the experimental trials: (i) failed calibration of the eye-tracker; (ii) technical problems and software failures; (iii) the infants' behavior could not contribute to complete the experimental trials due fussiness or audible crying. A total of 26 participants were excluded from the analyses, 13 of them after we applied the exclusion criteria filters, and the other 13 were reported by the experimenter as non-valid, either because they were not engaged in the task (e.g., less than 0.5 s of fixation to the screen at the pre-naming period) or because the parent interfered (e.g., attracted the baby's attention, moved their body, or adjusted the seat).

The measures of pupil size were selected carefully during each trial's windows of interest: pre-naming (baseline) window = last 100 ms before the auditory target word onset were presented, and post-naming window = 3000 ms, from target word onset to the end of the trial (see Figure 5).

Figure 5

Illustration of Our Trial Structure. The first 1500 ms with an Isoluminant Screen, followed by 1000 ms Long Silent Image of a Target Item. Then Our Baseline Window (Last 100 ms Before the Target Word is Named) And the Post-naming Window (3000 ms After Target Word Onset.



We baseline corrected the pupil sizes by subtracting the data obtained at the pre-naming period from the post-naming period. Then, as previously described at the design section, we computed the difference between the average baseline-corrected pupil sizes of matching and unrelated trials. Note that not all words had valid results in both conditions, for example, we could obtain the average baseline-corrected pupil size for one of our participants in the trial "apple", for matching condition, and yet not be the case for the unrelated condition. In such cases we excluded the whole trial from the data set, in order to strictly follow the experiment protocols announced in our pre-registered report (https://osf.io/6hzmc). Results from trials where infants did not look at the screen for at least a 50% of the trial's duration were also excluded. In total, we obtained the valid pupil data on 587 trials to our final sample of participants.

Subsequently, we separated our data to perform two analyses: the first included data from the words that the parents had reported as never used in the child's presence; and the second had these words excluded. Consequently, the analysis number two had less items per participant than the analysis number one.

Results

Dependent variable

We analyzed word comprehension by measuring pupil sizes, presenting images of eight highly familiar word referents in two different conditions: a matching label and a non-matching and contextually unrelated label. Interpreting that words that are understood for the infant display higher degrees of pupil dilation at the unrelated trials, indicating greater cognitive effort from the infant by trying to establish a connection between a known word and an unrelated label. Pupil sizes were compared to words reported as understood and not-understood in two analyses: first, in an analysis including the words reported as never heard for the infant, and second, we analyzed difference in pupil sizes, compared to understood and not-understood words, excluding the words reported as never heard for the infant.

Statistical analysis

For this study we opted for mixed-effect models that take into account variability in random effects of items and infants (e.g., participants and word combinations), using the packages lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) in R program for statistical computing (R team). There were preformed linear mixed model analysis with fit by maximum likelihood and coefficient tables including t-tests and p-values using Satterthwaites's methods for degrees of freedom and t-statistics. The parental reports (words reported as understood vs. not understood) were set as a fixed factor, and the total comprehension scores are set as covariates. The differences in pupil size between conditions were set as a reference level and the anova function was used to assess the model effects as in Kartushina & Mayor (2019)

Analyses including words reported as non-heard by the children

Our first analysis includes the words reported as never being used in the infants' presence by their parents. Our variables: (i) the parental reports and (ii) the baseline-corrected average pupil size differences, were selected correspondingly (per participant, per item) to have identically matching data. In consequence, parental reported data from one participant, was removed since we did not have valid pupil size measures in both conditions, thus, we excluded one word from our overall vocabulary size, being a total of 304 words, where 214 of them were reported as not understood and 90 as understood.

The baseline-corrected average pupil-size differences (across matching and unrelated conditions) were selected per participant (n = 64), per object (eight items) including words reported by the parents as never heard by their children, having a mean score M = -.00 (*min.* = -.58, *max.* = .72), whereas the mean score of the total comprehension scores reported from the parents is M = 56 (*min*. = 4, *max*. = 205).

Our results using mixed model effect models show that the degrees of pupil dilation were not significantly different in the presented conditions (p > .01). Furthermore, this analysis reveals no significant effects on the parental reports of words reported as understood, suggesting that there were no differences between average pupil sizes and parental reports, see Table 8 for details.

Formula: difference_conditions ~ parent_report + total_comprehension_scores + gender + (1 | id) + (1 | word).

Table 6

Maximum Log-likehood and Information Criteria

AIC	BIC	Log-likelihood	Deviance	*df*. Residuals
-250.8	-224.8	132.4	-264.8	297

Table 7

Scaled Residuals

Minimum	Q1	Median	Q3	Maximum
-3.76	-0.43	-0.01	0.43	4.47

Table 8

Results for the Linear Mixed Model Fit by Maximum Likelihood (t-test with Satterthwaite's Method)

Random effects	Variance		*SD*		
Id	1.84		4.30		
Word	1.38		1.17		
Residual	2.45		1.56		
Fixed effects	Estimate	*SE*	*df.*	t	*p*
Intercept	-1.18	1.87	3.04	-.63	.52
Parental reports	351	209	3.04	167	.09.
Total comprehension scores	-2.21	2.13	3.04	-1.03	.30

. $p < 0.1$ * $p < .05$ **$p < .05$ *** $p < .001$

As shown in Table 8, there were no significant main effects of the difference of pupil sizes in matching and unrelated conditions, thus we cannot say that we find evidence of word comprehension in this analysis. Neither the words reported as understood/not understood or the total comprehension scores were significant factors when included in the regression.

Table 9

Correlation of Fixed Effects

	Intercept	Parental reports	Comprehension scores

Intercept	-.08	-	-
Parental reports	-.55	-.34	-
gender	-.62	-.04	.03

Table 10

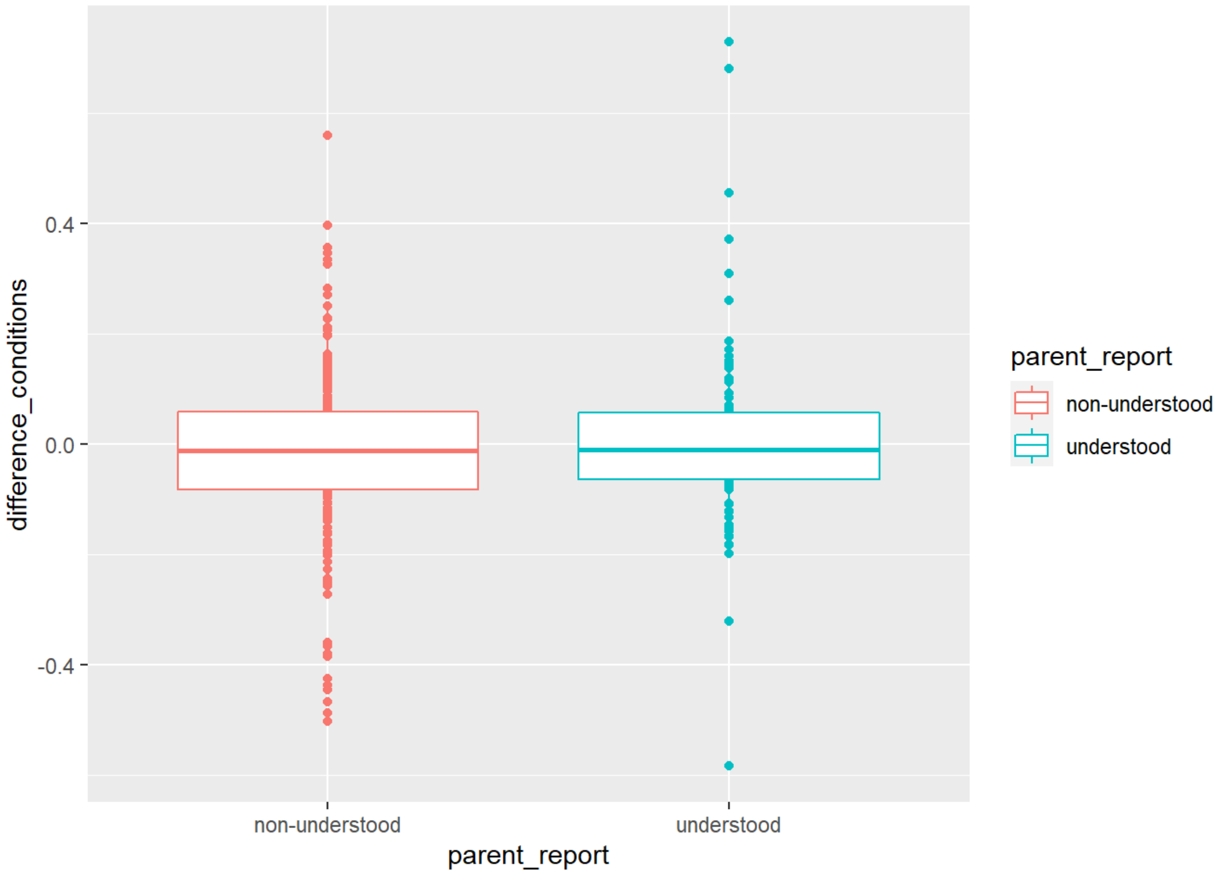
Type III Analysis of Variance with Satterthwaite's Method

	Sum of squares	Mean sq	*df.Num*	*df.Den*	*F*	*p*
Parental reports	.06	.06	1	304	2.82	.09.
Total comprehension scores	.02	.02	1	304	1.07	.30
Gender	.00	.00	1	304	.15	.69

. $p < 0.1$ * $p < .05$ ** $p < .05$ *** $p < .001$

Figure 5

Illustration of the difference of pupil sizes between conditions, in comparison to the parental reports in analysis that include non-heard words.



Follow up analyses with generalized linear mixed models fit by maximum likelihood (Laplace approximation) were performed to approach parents' reports per item with comprehension as a predictor, yet these show no significant difference from zero: Estimate = -.28, *SE* = .25, z = -1.12, *p* = .25.

Analyses excluding words reported as non-heard by the children

In line with previous investigations, we decided to perform an analysis excluding those words reported as never used in the infants' presence since they were born (Bergelson & Swingley, 2015; Houston-Price et al., 2007; Syrnyk & Meints, 2017) using the exact same formula than our analyses excluding the non-heard words. Yet this time, we analyzed les items due the exclusion of the word referents reported as never being said in the infants' presence, being a total of 283 words, where 193 of them are reported as not understood and 90 are reported as understood.

The mean score of our pupil size data selected for this analysis was *M* = -.01 (*min*. -.58, *max*. .72). Our statistical analysis showed tendentially significant effects at the understood section.

In this analysis, our results show a tendentially significant effect of parental reports, yet the effects of word comprehension are not significant. See Table 13 for details.

Table 11

Maximum Log-likehood and Information Criteria

AIC	BIC	Log-likelihood	Deviance	*df*. Residuals
-219	-193.6	116.5	-233	270

Table 12

Scaled Residuals

Minimum	Q1	Median	Q3	Maximum
-3.65	-.40	.00	.43	4.36

Table 13

Results for the linear mixed model fit by maximum likelihood

Random effects	Variance		*SD*		
Id	.00		.00		
Word	.00		.00		
Residual	.02		.15		
Fixed effects	Estimate	*SE*	*df.*	t	*p*
Intercept	-2.22	2.05	6.30	-1.08	.28
Parental reports	4.05	2.23	4.59	1.81	.07.

Total comprehension scores	-2.66	2.33	2.69	-1.14	.25

. $p < 0.1$ * $p < .05$ ** $p < .05$ *** $p < .001$

As shown in table 13, the results of mixed-effects regression analysis reveal a marginal effect of parental reports ($p = .07$), suggesting that there was a tendentially significant difference in pupil sizes for words reported as understood in comparison with words reported as not understood. All other effects are not significant.

Table 14

Correlation of Fixed Effects

	Intercept	Parental reports	Total comprehension scores
Intercept	-.08	-	-
Parental reports	-.53	-.38	-
gender	-.63	-.03	.05

Table 15

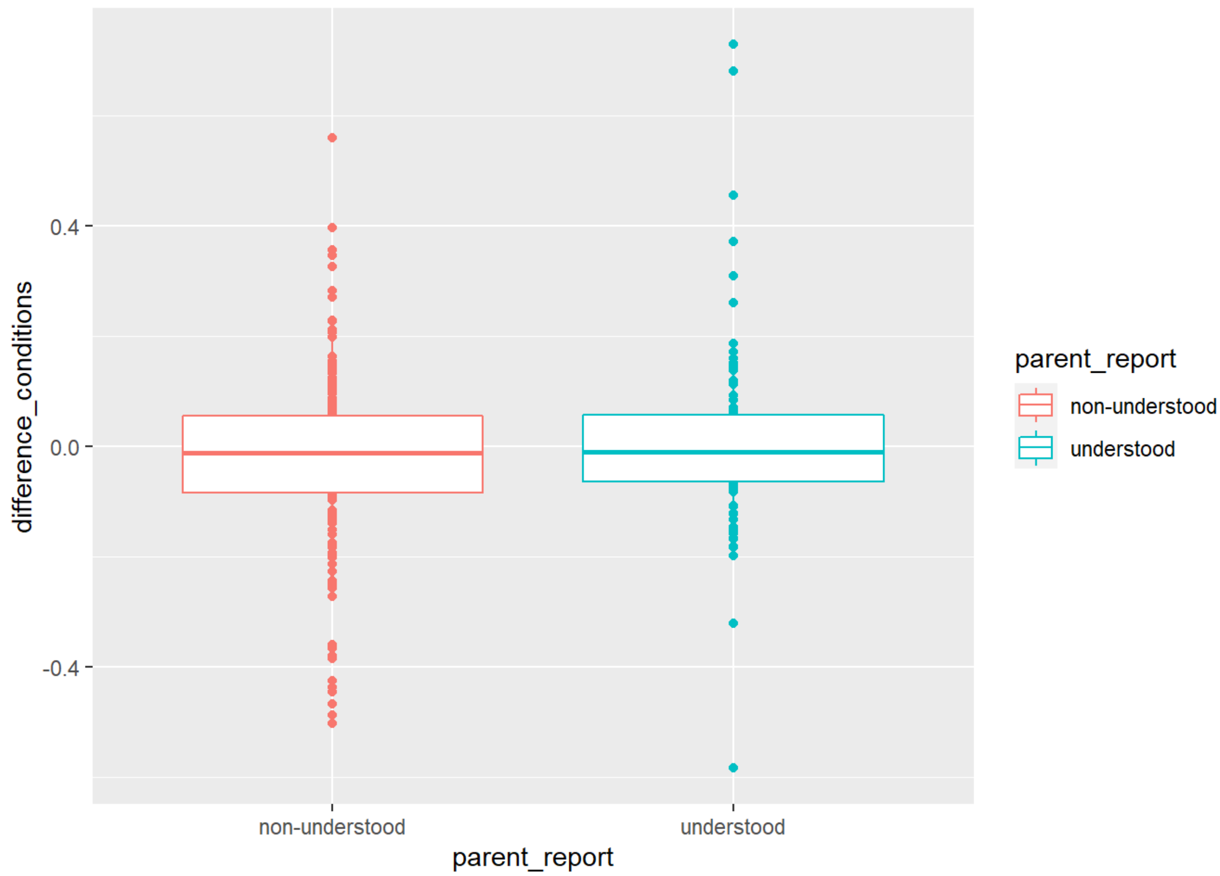
Type III Analysis of Variance with Satterthwaite's Method

	Sum of squares	Mean sq.	*df.Num*	*df.Den*	*F*	*p*
Parental reports	.08	.08	1	45.9	3.29	.07.
Total comprehension scores	.03	.03	1	269	1.30	.25
Gender	.02	.02	1	274	.83	.36

. $p < 0.1$ * $p < .05$ ** $p < .05$ *** $p < .001$

Figure 6

Illustration of the difference of pupil sizes between conditions, in comparison to the parental reports in analysis that exclude non-heard words, where a minimal effect is observed.



As in our first analysis including unknown words, we performed a follow up analysis with generalized linear mixed models fit by maximum likelihood (Laplace approximation), yet none of these effects were significant either: Estimate = -.29, *SE* = .26, z = -1.09, *p* = .27.

Discussion

The main goals of this thesis were: first, to contribute to the limited scientific data of word comprehension on Norwegian infants, and second, to analyze the relationship between parental reports on their infants' early word knowledge and the outcomes from a direct measure of word comprehension, as in Syrnyk & Meints (2017). In brief, the present findings reveal (i) no alignment between our results and the ones obtained by Syrnyk & Meints (2017), where English speaking infants around the age of 9- months show word comprehension for those items reported by their parents as understood and not understood, suggesting that parents are good assessors of their infants' lexical development; and (ii) we found non-significant effects of pupil size differences between matching and unrelated conditions, suggesting that our population (*n* = 64) of 8- to 12-month-old Norwegian infants did not show evidence of word comprehension for familiar or unfamiliar words.

Summed up, in our analyses, we explored infants' word comprehension in a pupillometry study, measuring the differences in baseline-corrected pupil size averages under two conditions (matching and unrelated objects) and found no solid evidence of word comprehension, in consequence, the effects of parental assessment of their infants' word recognition in this study cannot be expected to be robust either.

Reliability of Parental Reports Assessing Infants' Word Comprehension

According to our results, the evidence of Norwegian parents being reliable assessors or their infants' word knowledge at ages 9- 12-months is not robust and is in odd with our hypothesis, however our results reveal marginally significant effects of words reported as understood in our analysis when excluding non-familiar words for the infants. These findings suggest that parents' predictions skills might be accurate under other methodological conditions (e.g., at later ages and with a wider range of familiar word referents).

Another possible interpretation of these results could be in line with previous studies assessing early word comprehension, which have revealed that infants rely in diverse lexical cues (e.g., semantic in Bergelson & Swingley, 2017; contextual in Arias-Trejo & Plunkett, 2010; and frecuency in Kartushina & Mayor, 2019), suggesting that infants stablish a net of word – object associations to facilitate word comprehension. These associations (or word mappings) have been observed in infants around six months of age (recognizing familiar items as e.g., banana, bottle, hand in Bergelson & Swingley, 2012) and have been suggested to influence the parents' intuitions to predict word understanding in their infants. For example, in a study with older Norwegian children (18- 20-months) Lo and colleagues (2021), found agreement between parental reports and their toddlers' word knowledge in word comprehension trials, where words were cued by rich context. These results reveal that parents' predictions of their children's word comprehension are not exclusively based on the child's ability to comprehend words per se, their reports are also based on the child's ability to make use of contextually rich associations.

So in one hand, although our results did not show significant effects of word comprehension, we observe tendentially significant effects when including the parental reports of known words in the analyses *(p* = .07) and can thus speculate that parents' reports relied on words cued by frequency of occurrence. These assumptions support previous findings of parental reports on their infants' word comprehension, sustaining that parents' predictions are more accurate when they assess for words that are known for the child (Bergelson & Swingley, 2015; Houston-Price et al., 2007; Kartushina & Mayor, 2022; Syrnyk & Meints, 2017). However, in the other hand, assumptions should be taken carefully, since we did not test for contextually related items.

Further studies of word comprehension and parental reports in Norwegian infants

In this study, although the words known to the infants were not selected from individualized word lists (which deviates from Syrnyk & Meints, 2017's study) and instead, we measured comprehension with items that are rated as highly familiar to infants between 9 and 12-months by the standards indexed at WordBank, the median vocabulary sizes reported from our participants: M = 48, were very similar to the median vocabulary sizes previously reported in Norwegian infants between the same ages: M = 47 (retrieved from http://wordbank.stanford.edu/, Frank et al., 2017). Which suggests that the vocabulary size obtained from our participants represents the normative data for vocabulary size of Norwegian infants in ages from 9- 12-months. We selected our target word stimuli under the constructs of food/kitchen and body/cleaning routine objects, in line with previous research (Bergelson & Swingley, 2012; Kartushina & Mayor, 2019) hence, we were not able to select word stimuli from very familiar items, whether reported from the parents or from the very familiar words submitted in the CDI's normative data. Alternately, we ask the parents rate word familiarity of some highly familiar objects that matched our criteria. However, further pupil dilation studies measuring early word comprehension with very familiar items in an individualized fashion, are highly suggested. First, to explore word comprehension, as it has been documented to have its emergence in ages around nine months for Norwegian children (Kartushina & Mayor, 2019); and second, to investigate parents' reliability assessing infants' word comprehension in ages below one year, as intended in the present study.

Direct Measures and Outcomes of Word Comprehension

The results from our experiment did not show statistically significant effects in the difference of pupil sizes across conditions, hence, we find no evidence of infants' displaying word recognition in our trials. The responses on pupillary sizes between our two analyses (known vs unknown words) were different from each other, yet none of them were significant. According to these results, our findings of word comprehension can indicate that our population's word representations are brittle and still rely on the semantic content and the context and/or activity in which they were uttered (Tamis-Lemonda et al., 2018).

Diverse studies have explored early word representations with direct measures in English speaking infants, before their first year of life. Some of the authors behind these studies described in their results that infants have abilities to successfully associate a diverse range of labels to its appropriate word referent (Bergelson & Aslin, 2017; Bergelson & Swingley; 2012; Syrnyk & Meints; 2017; Tincoff & Jusczyk; 1999), as early as at ages of six months of age. However, as for Norwegian children the first signs of emergence of word comprehension occurs a little later, between eight and nine months of age (Kartushina & Mayor, 2019). In the present study we attempted to corroborate the findings of word comprehension with analyses of pupil dilation in infants at ages from eight to 12-months. Note that in their study Kartushina and Mayor (2019) explain that word representations in Norwegian infants between eight and nine months of age are coarse, and infants failed to disambiguate between objects without extra linguistic cues. Hence, is no surprising for us to speculate that word representations remain coarse in our population, since we tested for word referents matching the picture of a depicted object, versus semantically and contextually unrelated labels, hence, no cues for the infants to rely on. However, pupil dilation analyses could be a suitable option to measure word comprehension in Norwegian children at later ages, e.g., 13- 15-months, as word representations in Norwegian children seem to be more robust after the age of one year (Kartushina & Mayor, 2022; Lo et al., 2019).

Conclusions

According to our results, Norwegian parents do not seem to be good assessors of their infants' early word knowledge, the past findings reinforce the suggestions of using supplemental measures in the assessment of infants' word knowledge, rather than the use of parental reports as a sole measure, as mentioned by diverse authors (Fenson et al., 1993; Houston-Price et al., 2007; Kartushina & Mayor; 2019; Lo et al., 2021; Syrnyk & Meints, 2017), particularly when applied to children under 12 months of age. However, the marginally significant effect found in our analysis, where the words reported as unknown were excluded might suggest two things: first, Norwegian parents might be good assessors of their infants' word comprehension, yet at such early ages, word representations remain weak and under these circumstances predicting word understanding is a complex task for the parents; and second, that measures of pupil dilation can be used to explore word comprehension and its relationship with parental of their child's word comprehension, but needs to be tested in older children and preferably with more familiar items.

References

Arias-Trejo, N., & Plunkett, K. (2010). The effects of perceptual similarity and category membership on early word-referent identification. *J Exp Child Psychol*, *105*(1), 63-80. https://doi.org/10.1016/j.jecp.2009.10.002

Arriaga, R. I., Fenson, L., Cronan, T., & Pethick, S. J. (1998). Scores on the MacArthur Communicative Development Inventory of children from low and middle-income families. *Applied Psycholinguistics*, *19*(2), 209-223.

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

Bauer, D. J., Goldfield, B. A., & Reznick, J. S. (2002). Alternative approaches to analyzing individual differences in the rate of early vocabulary development. *Applied Psycholinguistics*, *23*(3), 313-335. https://doi.org/10.1017/S0142716402003016

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, *91*(2), 276-292. https://doi.org/10.1037/0033-2909.91.2.276

Beatty J, Lucero-Wagoner B (2000) The pupillary system. In: Caccioppo J, Tassinary LG, Berntson, G editors. The Handbook of Psychophysiology. Hillsdale: Cambridge University Press.

Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proc Natl Acad Sci U S A*, *114*(49), 12916-12921. https://doi.org/10.1073/pnas.1712966114

Bergelson, E., & Aslin, R. N. (2017b). Semantic Specificity in One-Year-Olds' Word Comprehension. *Language Learning and Development*, *13*(4), 481–501. https://doi.org/10.1080/15475441.2017.1324308

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proc Natl Acad Sci U S A*, *109*(9), 3253-3258. https://doi.org/10.1073/pnas.1113380109

Bergelson, E., & Swingley, D. (2015). Early Word Comprehension in Infants: Replication and Extension. Language Learning and Development, 11(4), 369–380. https://doi.org/10.1080/15475441.2014.979387

Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition*, *127*(3), 391-397. https://doi.org/10.1016/j.cognition.2013.02.011

Bloom, L. (1970). *Language development: form and function in emerging grammars* (Vol. 59). MIT Press.

Bloom, P. (1993). *Language acquisition: core readings*. Harvester Wheatsheaf.

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer software]. http://www.praat.org/

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annu Rev Psychol*, *53*(1), 371-399. https://doi.org/10.1146/annurev.psych.53.100901.135233

Camaioni, L., Castelli, M. C., Longobardi, E., & Volterra, V. (1991). A parent report instrument for early language assessment. *First language*, *11*(33), 345-358. https://doi.org/10.1177/014272379101103303

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, *25*, 69–91. https://doi.org/10.1016/j.dcn.2016.11.001

Eckstein, M. K., Starr, A., & Bunge, S. A. (2019). How the inference of hierarchical rules unfolds over time. *Cognition*, 185, 151–162. https://doi.org/10.1016/j.cognition.2019.01.009

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement Properties of the MacArthur Communicative Development Inventories at Ages One and Two Years. *Child Dev*, *71*(2), 310-322. https://doi.org/10.1111/1467-8624.00146

Egger, J., Rowland, C. & Bergmann, C. (2020). Improving the robustness of infant lexical processing speed measures. *Behavior Research Methods*, *52*(5), 2188–2201. https://doi.org/10.3758/s13428-020-01385-5

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in Early Communicative Development. *Monogr Soc Res Child Dev*, *59*(5), i-185. https://doi.org/10.2307/1166093

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*, *21*(1), 95-116. https://doi.org/10.1017/S0142716400001053

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Dev Sci*, *16*(2), 234-248. https://doi.org/10.1111/desc.12019

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. *Developmental psycholinguistics: On-line methods in children's language processing*, *44*, 97.

Fló, A. (2021). Evidence of ordinal position encoding of sequences extracted from continuous speech. Cognition, 213, 104646–104646. https://doi.org/10.1016/j.cognition.2021.104646

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: an open repository for developmental vocabulary data. *J. Child Lang*, *44*(3), 677-694. https://doi.org/10.1017/S0305000916000209

Friedrich, M., & Friederici, A. D. (2011). Word Learning in 6-Month-Olds: Fast Encoding–Weak Retention. *Journal of Cognitive Neuroscience*, *23*(11), 3228-3240. https://doi.org/10.1162/jocn_a_00002

Geangu, E., Hauf, P., Bhardwaj, R., & Bentz, W. (2011). Infant pupil diameter changes in response to others' positive and negative emotions. *PLoS One*, *6*(11), e27132-e27132. https://doi.org/10.1371/journal.pone.0027132

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annu Rev Psychol*, *61*(1), 191-218. https://doi.org/10.1146/annurev.psych.093008.100408

Gervain, J. (2015). Plasticity in early language acquisition: the effects of prenatal and early childhood experience. Current Opinion in Neurobiology, 35, 13–20. https://doi.org/10.1016/j.conb.2015.05.004

Ghio, M., Cara, C., & Tettamanti, M. (2021). The prenatal brain readiness for speech processing: A review on foetal development of auditory and primordial language networks. Neuroscience and Biobehavioral Reviews, 128, 709–719. https://doi.org/10.1016/j.neubiorev.2021.07.009

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of child language*, *14*(1), 23-45. https://doi:10.1017/s030500090001271x

Gredebäck, G., & Melinder, A. (2010). Infants' understanding of everyday social interactions: A dual process account. *Cognition*, *114*(2), 197-206. https://doi.org/10.1016/j.cognition.2009.09.004

Hayes, T. R., & Petrov, A. A. (2015). Mapping and correcting the influence of gaze position on pupil size measurements. Behavior Research Methods, 48(2), 510–527. https://doi.org/10.3758/s13428-015-0588-x

Hepach, R., & Westermann, G. (2016). Pupillometry in Infancy Research. *Journal of cognition and development*, *17*(3), 359-377. https://doi.org/10.1080/15248372.2015.1135801

Hess, E. H., & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science*, *143*(3611), 1190-1192. https://doi.org/10.1126/science.143.3611.1190

Hochmann, J.-R. (2013). Pupillometry in six-month-old infants. Proceedings of the 37th annual conference on language development.

Hochmann, J.-R. & Papeo, L. (2014). The Invariance Problem in Infancy. Psychological Science, 25(11), 2038–2046. https://doi.org/10.1177/0956797614547918

Hoff, E. (2003). The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Dev*, *74*(5), 1368-1378. https://doi.org/10.1111/1467-8624.00612

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *J. Child Lang*, *34*(4), 701-724. https://doi.org/10.1017/S0305000907008124

How do Tobii Eye Trackers work? *Learn more with Tobii Pro*. (2015). https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/how-do-tobii-eye-trackers-work/

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

Jusczyk, P. W., & Hohne, E. A. (1997). Infants' Memory for Spoken Words. *Science*, *277*(5334), 1984-1986. https://doi.org/10.1126/science.277.5334.1984

Kartushina, N., & Mayor, J. (2022). Coping with dialects from birth: Role of variability on infants' early language development. Insights from Norwegian dialects. Developmental Science, e13264–e13264. https://doi.org/10.1111/desc.13264

Kartushina, N., & Mayor, J. (2019). Word knowledge in six- To nine-month-old Norwegian infants? Not without additional frequency cues. *Royal Society open science*, *6*(9), 180711-180711. https://doi.org/10.1098/rsos.180711

Lo, C. H., Rosslund, A., Chai, J. H., Mayor, J., & Kartushina, N. (2021). Tablet assessment of word comprehension reveals coarse word representations in 18–20-month-old toddlers. *Infancy*, *26*(4), 596–616. https://doi.org/10.1111/infa.12401

MacWhinney, B. (2019). Understanding spoken language through TalkBank. *Behav. Res. Methods*, *51*(4), 1919-1927. https://doi.org/10.3758/s13428-018-1174-9

Mathôt, S., Elle-Van Heusden, J. F., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. Behavior Research Methods, 50(1), 94–106. https://doi.org/10.3758/s13428-017-1007-2

Mayor, J., & Mani, N. (2018). A short version of the MacArthur–Bates Communicative Development Inventories with high validity. Behavior Research Methods, 51(5), 2248–2255. https://doi.org/10.3758/s13428-018-1146-0

McDougal, & Gamlin, P. D. (2011). Autonomic Control of the Eye. *Compr*. *Physiol*. (2015) 5:439–73 https://doi.org/10.1002/cphy.c140014

McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, *95*(2), B15-B26. https://doi.org/10.1016/j.cognition.2004.07.005

Miller, W. R. (1972). Aaron Bar-Adon and Werner Leopold (eds), Child language, A book of readings. Englewood Cliffs, New Jersey: Prentice-Hall, 1971. *Language in society*, *1*(2), 285-290. https://doi.org/10.1017/S0047404500000609

Napieralski, P. & Rynkiewicz, F. (2019). Modeling Human Pupil Dilation to Decouple the Pupillary Light Reflex. *Open Physics*, *17*(1), 458–467. https://doi.org/10.1515/phys-2019-0047

Nelson, D. G. K., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant behavior and development*, *18*(1), 111-116.

Parise, E., & Csibra, G. (2012). Electrophysiological Evidence for the Understanding of Maternal Speech by 9-Month-Old Infants. *Psychol Sci*, *23*(7), 728-733. https://doi.org/10.1177/0956797612438734

Reznick, J. S., & Goldsmith, L. (1989). A multiple form word production checklist for assessing early language. *Journal of child language*, *16*(1), 91-100.

Sachse, S., & Von Suchodoletz, W. (2008). Early identification of language delay by direct language assessment or parent report? *J Dev Behav Pediatr*, *29*(1), 34-41. https://doi.org/10.1097/DBP.0b013e318146902a

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews. Neuroscience*, *10*(3), 211–223. https://doi.org/10.1038/nrn2573

Schafer, G. (2005). Infants can learn decontextualized words before their first birthday. *Child Dev*, *76*(1), 87-96. https://doi.org/10.1111/j.1467-8624.2005.00831.x

Schmidtke, J. (2018). PUPILLOMETRY IN LINGUISTIC RESEARCH: AN INTRODUCTION AND REVIEW FOR SECOND LANGUAGE RESEARCHERS. *Studies in second language acquisition*, *40*(3), 529-549. https://doi.org/10.1017/S0272263117000195

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

Styles, S., & Plunkett, K. I. M. (2008). What is ‘word understanding’ for the parent of a one-year-old? Matching the difficulty of a lexical comprehension task to parental CDI report. *J. Child Lang*, *36*(4), 895-908. https://doi.org/10.1017/S0305000908009264

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Dev Sci*, *8*(5), 432-443. https://doi.org/10.1111/j.1467-7687.2005.00432.x

Swingley, D. (2011). The looking-while-listening procedure. *Guide to Research Methods in Child Language*.

Syrnyk, C., & Meints, K. (2017). Bye-bye mummy – Word comprehension in 9-month-old infants. *Br J Dev Psychol*, *35*(2), 202-217. https://doi.org/10.1111/bjdp.12157

Tamási, K., McKean, C., Gafos, A., Fritzsche, T., & Höhle, B. (2017). Pupillometry registers toddlers' sensitivity to degrees of mispronunciation. *J Exp Child Psychol*, *153*, 140-148. https://doi.org/10.1016/j.jecp.2016.07.014

Tamasi, K., McKean, C., Gafos, A., & Hohle, B. (2019). Children's gradient sensitivity to phonological mismatch: considering the dynamics of looking behavior and pupil dilation. *J Child Lang*, *46*(1), 1-23. https://doi.org/10.1017/S0305000918000259

Tamis-LeMonda, C., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019). Routine Language: Speech Directed to Infants During Home Activities. Child Development, 90(6), 2135–2152. https://doi.org/10.1111/cdev.13089

Tincoff, & Jusczyk, P. W. (1999). Some Beginnings of Word Comprehension in 6-Month-Olds. Psychological Science, 10(2), 172–175. https://doi.org/10.1111/1467-9280.00127

Tomasello, M., & Mervis, C. B. (1994). THE INSTRUMENT IS GREAT, BUT MEASURING COMPREHENSION IS STILL A PROBLEM. *Monographs of the Society for Research in Child Development*, *59*(5), 174-179. https://doi.org/10.1111/j.1540-5834.1994.tb00186.x

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychon Bull Rev*, *25*(6), 2005-2015. https://doi.org/10.3758/s13423-018-1432-y

Yoder, P. J., Warren, S. F., & Biggar, H. A. (1997). Stability of maternal reports of lexical comprehension in very young children with developmental delays. *American Journal of Speech-Language Pathology*, *6*(1), 59-64.

Zele, A. J. & Gamlin, P. D. (2020). Editorial: The Pupil: Behavior, Anatomy, Physiology and Clinical Biomarkers. *Front. Neurol. 11:211*. https://doi.org/10.3389/fneur.2020.00211

Zhang, F., Jaffe-Dax, S., Wilson, R. C., & Emberson, L. L. (2019). Prediction in infants and adults: A pupillometry study. *Dev Sci*, *22*(4), e12780-n/a. https://doi.org/10.1111/desc.12780

APPENDIX A

Stimuli Selection by Higher Frequency Rank on CDI WG and Childes

	Food/kitchen related targets				Body/Cleaning_bath-related targets			
6	banana	bread	apple	cookie	foot	belly	hair	leg
Norwegian	(En) banan	(Et) brød	(Et) eple	(En) kjeks	(En) fot	(En) mage	(Et) hår	(Et) bein
Transcription	bɑnɑ:n	brø:	eplə	çeks	fu:t	mɑ:gə	ho:r	bæjn
Word frequency	107	26	40	20	11	6	47	16
CDI 11 months	46	32	19	17	23	19	17	7
CS	Se! En banan!					Se! En mage!		
		Se på brødet!						Se på beinet!
			Her er eplet!	Her er kjeksen!				
					Oi, en fot!		Oi, et hår!	
	Contextually-related words							
English	spoon	cup	chair	table	diaper	toothbrush	bathtub	sock
Norwegian	(En) skje	(En) kopp	(En) stol	(Et) bord	(En) bleie	(En) tannbørste	(Et) badekar	(En) sokk
Transcription	ʃe:	kɔp	sto:l	bɔrd	blæjə	tɑnbœʃtə	bɑ:dəkɑr	so´k
Word frequency	6	17	40	4	34	8	2	19
CDI 11 months	20	25	14	14	46	27	25	7
CS	Se! En skje!					Se! En tannbørste!		
		Se på koppen!						Se på sokken!
			Her er stolen!	Her er bordet!				
					Oi, en bleie!		Oi, et badekar!	
	Random words							
English	doll	telephone	pants	picture	pacifier	stroller	light (object)	box
Norwegian	(En) dukke	(En) telefon	(En) bukse	(Et) bilde	(En) smokk	(En) barnevogn	(Et) lys	(En) eske
Transcription	du`k:ə	teləfo:´n	buksə	bi´ldə	smok:	ba:`rnəvåŋn	ly:s	e`skə
Word frequency	36	20	18	57	6	29	13	2
CDI 11 months	20	27	14	13	43	28	29	8
CS	Se! En dukke!					Se! En barnevogn!		
		Se på telefonen!						Se på esken!
			Her er buksen!	Her er bildet!				
					Oi, en smokk!		Oi, et lys!	

APPENDIX B

Audio Stimuli: Sentences' Length and Target Word Onset

	A	B	C	D
1	File name	audio total length (ms)	target word start (ms)	Time difference
2	2.Banana	2.069683	1.315269	0.754414
3	3.Spoon	2.103265	1.251187	0.852078
4	4.Doll	2.230363	1.559031	0.671332
5	5.Bread	1.280249	0.667256	0.612993
6	6.Cup	1.25932	0.66617	0.59315
7	7.Telephone	1.503469	0.611346	0.892123
8	8.Apple	1.09093	0.459752	0.631178
9	9.Chair	1.284218	0.506625	0.777593
10	10.Pants	1.319887	0.642812	0.677075
11	11.Cookie	1.10263	0.402091	0.700539
12	12.Table	0.994422	0.44485	0.549572
13	13.Picture	1.054127	0.46339	0.590737
14	14.Foot	1.958844	1.04681	0.912034
15	15.Diaper	1.49263	0.739603	0.753027
16	16.Pacifier	1.803651	1.013905	0.789746
17	17.Belly	2.68288	1.851197	0.831683
18	18.Thootbrush	2.27127	1.326916	0.944354
19	19.Stroller	2.052812	1.083232	0.96958
20	20.Hair	1.778141	1.058334	0.719807
21	21.Bathtub	1.802766	0.762648	1.040118
22	22.Light	1.658481	0.893531	0.76495
23	23.Leg	1.05712	0.494831	0.562289
24	24.Sock	1.101882	0.507871	0.594011
25	25.Box	1.222766	0.618341	0.604425

APPENDIX C1: Visual Stimuli "Eple"

APPENDIX C2: Visual Stimuli "Banan"

APPENDIX C3: Visual Stimuli "Mage"

APPENDIX C4: Visual Stimuli "Brød"

APPENDIX C5: Visual Stimuli "Kjeks"

APPENDIX C6: Visual Stimuli “Fot”

APPENDIX C7: Visual Stimuli "Hår"

APPENDIX C8: Visual Stimuli "Bein"

APPENDIX D

Example of booking form with link to CDI-WG questionnaire

UiO : Psykologisk institutt
Det samfunnsvitenskapelige fakultet

babyling

Hei og velkommen til oss på Babyling!

Vi takker deg for interessen og setter pris på at dere vil bidra til prosjektet vårt om språkutvikling og ord-representasjoner hos spedbarn. **Det vi skal gjøre nå er a planlegge ditt besøk til babylaboratoriet**, som i alt vil ta omtrent 20 minutter.

Vi ber deg om å vennligst les gjennom vedlagte dokument: Informasjonsbrev_babyling. Der vil du finne mer informasjon om prosjektet og hva det innebærer å delta.

For å avtale et tidspunkt som er mest mulig tilpasset deres behov, kan du velge å møte opp innen følgende tidsrom: 15 februar-5 mars 2021, fra kl.09:00-17:00.

Gjerne svar på dette brevet ved å presisere dato og klokkeslett, så får du en bekreftelse på dette sammen med deltaker nr. deres.

Om Covid-19 og smitteverntiltak:

Det vil aldri være mer enn 2 personer til stede på laboratoriet mens dere gjennomfører testingen, dette for å holde dere trygge og sørge for at dere får et hyggelig opphold hos oss på laboratoriet.

Vi har også lagt ved dokumentet: Smittevern_babyling, der du kan finne alle våre retningslinjer for å hindre smittespredning og deres møte med personalet.

Til slutt vil vi be deg å fylle ut følgende skjema: https://nettskjema.no/a/169497. Dette kan du gjøre når det best passer deg, **før** deres besøk på laboratoriet.

Vi er nå på vent etter ditt svar om passende tidspunkt.

Deres bidrag til vitenskapen er stort og viktig for samfunnet! Det er noe å være kjempestolt av. *Tusen takk fra alle oss.*