The perceptual basis of meaning acquisition

Auditory associative word learning and the effect of object modality on word learning in infancy and adulthood

Dissertation
zur Erlangung des Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
im Fachbereich Humanwissenschaften der
Universität Osnabrück

UNIVERSITÄT OSNABRÜCK

vorgelegt
von

Samuel H. Cosper

Osnabrück, 2020

This is a cumulative doctoral dissertation that includes both published articles and manuscripts submitted for publication in peer-reviewed journals. The articles are included in Part II. In this online version, Chapter 5 has been redacted due to copyright issues.

Supervisor and first examiner: Prof. Dr. Jutta L. Müller

Supervisor and Co-examiner: Prof. Dr. Claudia Männel

Co-examiner: Prof. Dr. Peter König

Abstract

The world in which we live is filled with sensory experiences. Language provides us with a manner in which to communicate these experiences with one another. In order to partake in this communication, it is necessary to acquire labels for things we see, hear, feel, smell, and taste. Much is known about how we learn words for things we can see, but this bias in the literature leaves many open questions about words attributed to other modalities. This cumulative dissertation aims to close this gap by investigating how both 10- to-12-month old infants and adults map novel pseudowords onto environmental sounds in an auditory associative word learning task with the aim to explore how humans learn words for things that cannot be seen, such as *thunder*, *siren*, and, *lullaby*. Infants were found, via event-related potentials (ERPs), to be successful at auditory associative word learning, while adults are much stronger learners in multimodal audio-visual conditions. Across the lifespan, sensory modality was found to affect word learning differently in infants than in adults. Where infants benefitted from unimodal auditory word learning, adults were more successful in multimodal audio-visual paradigms. Furthermore, the modality of the object being labelled modulated the temporal onset and the topological distribution of the N400 ERP component of violated lexical-semantic expectation. Lastly, the temporal congruency of presented stimuli affected word learning in adults in an inverted manner to other forms of statistical learning. Word learning is sensitive to age, modality, and means of presentation, providing evidence for various intertwined learning mechanisms and bringing us a step closer towards understanding human linguistic cognition.

Abstract – German

Die Welt, in der wir leben, ist voller Sinneserfahrungen. Die Sprache gibt uns die Möglichkeit, diese Erfahrungen miteinander zu kommunizieren. Um an dieser Kommunikation teilnehmen zu können, ist es notwendig, Bezeichnungen für Dinge zu erwerben, die wir sehen, hören, fühlen, riechen und schmecken. Es ist viel darüber bekannt, wie wir Wörter für Dinge lernen, die wir sehen können. Allerdings lässt diese Voreingenommenheit in der Literatur viele Fragen dazu offen, wie die Bezeichnungen anderer Modalitäten erlernt werden. Mit dieser kumulativen Dissertation möchte ich diese Lücke schließen, indem ich untersuche, wie sowohl 10 bis 12 Monate alte Säuglinge als auch Erwachsene in einer auditiv-assoziativen Wortlernaufgabe neue Pseudowörter auf Umweltgeräusche abbilden. Das Ziel ist also, zu untersuchen, wie Menschen Wörter für Dinge lernen, die wir nicht sehen können, wie z.B. *Donner*, *Sirene* und *Wiegenlied*. Es zeigte sich, dass Säuglinge über ereigniskorrelierte Potentiale (EKPs) beim auditiv-assoziativen Wortlernen erfolgreich sind, während Erwachsene unter multimodalen audiovisuellen Bedingungen viel stärker lernen. Über die gesamte Lebensspanne hinweg wurde festgestellt, dass die sensorische Modalität das Wortlernen bei Säuglingen anders beeinflusst als bei Erwachsenen. Während Säuglinge vom unimodalen auditorischen Wortlernen profitierten, waren Erwachsene bei multimodalen audiovisuellen Paradigmen erfolgreicher. Darüber hinaus änderte die Modalität des zu benennenden Objekts den zeitlichen Beginn und die topologische Verteilung der N400-EKP-Komponente der abweichenden lexikalisch-semantischen Erwartung. Schließlich beeinflusste die zeitliche Kongruenz der präsentierten Stimuli das Wortlernen bei Erwachsenen in umgekehrter Weise zu anderen Formen des statistischen Lernens. Das Lernen von Wörtern reagiert sensibel auf Alter, Modalität und Darstellungsmittel, liefert Beweise für verschiedene miteinander verflochtene Lernmechanismen und bringt uns dem Verständnis der menschlichen Sprachkognition einen Schritt näher.

Table of Contents

PART III: DISCUSSION

Part I:
General Introduction

1 Auditory associative word learning

The world in which we live can be experienced through many different sensory modalities including, the visual, auditory, haptic, olfactory, and gustatory modalities. As social beings, we communicate these experiences with others by means of language. This means that we have words to describe objects and events in these modalities. In order to communicate these experiences, however, first the words for the specific objects and events must be acquired. Much is known about how we learn words for things we can see (e.g., Borgström, Torkildsen, & Lindgren, 2015; Friedrich & Friederici, 2008; Horst & Samuelson, 2008; Junge, Cutler, & Hagoort, 2012; Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006; Smith & Yu, 2008; Taxitari, Twomey, Westermann, & Mani, 2019; Torkildsen et al., 2008; Torkildsen, Syversen, Simonsen, Moen, & Lindgren, 2007; Werker, Cohen, Lloyd, Casasola, & Stager, 1998); yet, many of these other modalities are left largely untouched. Many studies have been conducted in recent years to attempt to categorize the lexical items of language in terms of their sensory modality, which indicate that the modality with the most words allocated to it, at least for Indo-European languages and Chinese, is the visual modality (Chedid et al., 2019; I.-H. Chen, Zhao, Long, Lu, & Huang, 2019; Lynott, Connell, Brysbaert, Brand, & Carney, 2020; Majid et al., 2018; Miklashevsky, 2018; Speed & Majid, 2017; Vergallito, Petilli, & Marelli, 2020; Winter, Perlman, & Majid, 2018). However, there is evidence which indicates that some languages in the world do focus more on other modalities and that the hierarchy of always having a visual dominance within the lexical items of a language may not be universal (Majid et al., 2018). The goal if this dissertation is a step towards closing this gap by specifically investigating how we learn words for auditory objects, such as *thunder*, *lullaby*, and *siren* in both infants and adults in an associative word learning task. By examining the abilities of both infants and adults to map novel pseudowords onto auditory objects, we begin to have a more defined idea of how words affiliated with various sensory modalities are processed and acquired across the lifespan.

Within the scope of this dissertation on auditory associative word learning, one of the crucial underlying mechanisms is associative word learning. Associative word learning is a type of associative learning process pertaining to the acquisition of new words. For this type of learning to occur, the learner requires repeated exposure to the co-occurrence of a word and its label, which in turn strengthen the connection between the two over time (e.g., McMurray, Horst, & Samuelson, 2012; Sloutsky, Yim, Yao, & Dennis, 2017). This type of learning is also related to statistical learning. The term statistical learning describes a cognitive process which allows learners to extrapolate regularities based on underlying patterns from stimuli they perceive either in a laboratory setting or a natural setting (cf. Aslin, 2017; Schiavo & Froemke, 2019; Thiessen, 2017). Although this type of pattern recognition is not unique to humans (for a review, see Schiavo & Froemke, 2019), the focus in this dissertation is on human infants and human adults, specifically in unimodal auditory and visual statistical learning as well as audio-visual multimodal statistical learning. Although infants can apply statistical learning to both auditory and visual stimuli, infants have different preferences between the two modalities. Where infants display a familiarization preference for learned visual stimuli, a novelty preference is found in the auditory modality (Emberson, Misyak, Schwade, Christiansen, & Goldstein, 2019). Adults also have a preference between the visual and auditory modalities, as can be seen in more accurate responses to auditory sequences as compared to visual sequences in immediate recall (Conway & Christiansen, 2005); however, adults express a visual processing preference as visual object short-term memory is stronger than auditory object short-term memory (Cohen, Horowitz, & Wolfe, 2009; Yuval-Greenberg & Deouell, 2007, 2009). When confronted with multimodal audio-visual stimuli presented with differing frequencies of changes in the stimuli, children attend to the changes in the auditory, indicating an auditory dominance, while adults attend to the changes in the visual modality, displaying a visual preference (cf. Lewkowicz, 1988a, 1988b; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003). As word learning has mostly been experimentally explored with labeling visual objects, many of these word learning studies are multimodal.

Given these preferences in modality as described above, the goal of this dissertation is to assess the abilities of infants and adults to map spoken pseudowords onto auditory objects in the form of real environmental sounds in a unimodal auditory associative word learning paradigm, measuring their learning by means of electroencephalography (EEG) and the event-related potential (ERP) technique. In order to achieve this goal, the current dissertation is broken down into three main parts: a general introduction, the experiments, and the discussion. In Part I, the general introduction, Chapter 2 aims to provide a background for sensory processing and statistical learning abilities in both infants and adults. Here, the focus is on unimodal auditory processing and learning, unimodal visual processing and learning, as well as multimodal audio-visual processing and learning. In Chapter 3, word learning is presented. This chapter consists of the phenomenon of word learning, the abilities of humans to acquire new words, as well as the mechanisms behind word learning. Additionally, the neurophysiological method of using ERPs to measure word learning is presented in detail. Chapter 4 presents the infant and adult experiments as well as their hypotheses within the papers presented and between the age groups. Part II, the experiments, consists of two scientific papers submitted for peer-review that describe the experiments and their findings in detail. Chapter 5 is the experimental paper for infant auditory associative word learning. Chapter 6 presents a paper consisting of a 2x2 experimental design investigating auditory associative word learning as compared to multimodal audio-visual word learning in two temporal presentations in adults. Part III, the discussion, consists of Chapter 7, which provides a brief overview of the findings within the two papers and discusses how auditory associative word learning changes across the lifespan as well as the implications that this dissertation brings to word learning in general.

2 Sensory processing and development

In order to experience the world around us, we humans need to process information from various sources. In this chapter, the focus is on those types of information that stem from our natural environment including visual, auditory, gustatory, haptic, and olfactory; in other words, information processed through bodily experience. For the purpose of this dissertation, only two of these perceptual modalities are taken into consideration in greater detail – the visual and the auditory modality. Each is considered in unimodal processing as well as together in multimodal processing in two stages in life, focusing on healthy infant perceptual processing as well as in adulthood.

2.1 Sensory processing general

Throughout the human lifecycle, the ability to process perceptual information is of great importance. In typically developing infants, processing auditory information supports many functions one of which is the development of language (cf. Preissl, Lowery, & Eswaran, 2005), whereas processing in the visual modality is equally important and supports, for example, visuo-motor control of action (Braddick & Atkinson, 2011). These abilities develop throughout gestation and continue to further mature during infancy, and are still important processes in adulthood. However, there is an abundance of evidence that the processing of auditory and visual stimuli is subject to change across the lifespan (e.g. Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003; Thiessen, 2010).

This chapter focuses on these changes in sensory processing specifically in which infants and children focus on and prefer auditory stimuli over visual stimuli (e.g., Lewkowicz, 1988a, 1988b; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003) whereas adults tend to focus on and prefer visual stimuli (e.g., Robinson & Sloutsky, 2004). In light of this change in preference, Thiessen (2010) outlined two constraints as it pertains to statistical learning in the visual and auditory modalities – the developmental constraint and the stimulus constraint. The developmental constraint states that identical input does not always lead to identical output in learning between infants and adults. To better understand this claim, infant and adult sensory processing is discussed separately in this chapter. Thiessen's second constraint, the stimulus constraint, is related to the fact that adults recognize, for example, word-form as well as the relationship between auditory-verbal stimuli, i.e., words, and objects, but they have difficulties forming similar relationships between auditory-nonverbal stimuli, i.e., tonal sequences, and objects. To understand this constraint, both infant and adult subchapters focus on auditory and visual processing separately, whereas both auditory-verbal and auditory-nonverbal stimuli are addressed in the auditory processing subsections.

2.2 Sensory processing in infants

The ability to process perceptual information is vital for human development. As such, these processes begin in utero and develop over the course of gestation and continue to do so after birth. This can be seen in vital sensory processing abilities that can be measured by behavioral responses during fetal development, which, for example, can be measured using fetal magnetoencephalography (fMEG) (cf. Eswaran, Lowery, Wilson, Murphy, & Preissl, 2004; Preissl, Lowery, & Eswaran, 2005) as well as changes in heart-rate (for a review, see Lecanuet & Schaal, 1996). These abilities begin early on in the gestational timeline, which have been reported for reflex response to touch at 7 weeks gestation, response to light around 23 weeks gestation, and response to vibroacoustic stimulation at 24 weeks gestation (for a review, see Herschkowitz, 1988). Of course, these sensory processing abilities are not fully developed at the time of birth and require further maturation (e.g., Banks & Salapatek, 1983). Although the development of these abilities begin in utero, the main focus in this dissertation are the sensory processing capabilities and infants' ability to apply statistical learning after birth, particularly in the first year of life.

During development, infants are constantly experiencing high levels of sensory input, much of which could be novel, be it new sounds, sights, smells, textures, or even tastes. However, most of these experiences are also multimodal in nature. Tastes are accompanied by smells, televisions/tablets/mobile phones paying videos are accompanied by sound, and even textures are accompanied by the visual aspects of the object, to name a few examples. With all of this co-occurring information, one of the main questions is in these novel situations, what is the direct object of the infant's focus? If two modalities are both integral parts of a single object or action, what is more salient to the infant? In the remainder of this chapter, this question will be addressed based on the example of visual and auditory stimuli, looking at both in isolation as well as in co-occurring situations.

In infancy, there is evidence for an auditory preference also referred to as an auditory dominance (cf. Lewkowicz, 1988a, 1988b; Sloutsky & Napolitano, 2003). The term, however, may be misleading. Auditory dominance in this area refers to infants and children paying more attention to auditory stimuli than to visual stimuli when both are presented together. The auditory dominance in infancy is derived from audio-visual events, where young infants (6- and 10-month.olds) focus more on changes to the auditory stimuli than to changes in the visual stimuli (cf. Lewkowicz, 1988a, 1988b). This is described in greater detail in Chapter 2.2.1. In addition to the auditory dominance, statistical learning in unimodal and multimodal situations is addressed. Findings by Saffran (2002) have addressed constraints on statistical learning in auditory and visual modalities, which state that statistical learning in the visual modality benefits from simultaneous presentation, whereas statistical learning in the auditory modality benefits from sequential presentation (Saffran, 2002). These investigations into the sensory processing abilities and preferences of infants in the visual modality, the auditory modality, and multimodal audio-visual processing allows for comparisons to be made between modalities as well as across developmental stages.

2.2.1 Auditory Processing

In terms of human development, auditory processing begins even before birth with fetuses responding to sound sometime between the 24th and 27th week of gestation (Herschkowitz, 1988; Starr, Amlie, Martin, & Sanders, 1977; Weitzman & Graziani, 1968). This can be expanded upon by evidence indicating that a fetus in the womb shows, via heart-rate, a preference for the voice of the mother (Kisilevsky et al., 2003). Although sensory processing has its biological beginning in utero, the main focus in this Chapter is on the development of sensory processing after birth.

Auditory sensitivity is also found in neonates, via sucking rate, in their preference for the mother's voice (DeCasper & Fifer, 1980; DeCasper & Spence, 1986; Panneton & DeCasper, 1986), for stories heard whilst in the womb (DeCasper & Spence, 1986), and for the rhythmic pattern and prosody of their native language (Demany, 1982; Demany, McKenzie, & Vurpillot, 1977). Furthermore, neonates are able to distinguish between languages of different rhythmic classes; however, they are not able to distinguish languages within a rhythmic class (Nazzi, Bertoncini, & Mehler, 1998). Over the course of infancy, auditory preference for the native language further specializes over time to the point that 5-month-olds, found via head-turning, are not only capable of distinguishing their native language from other languages in the same native rhythmic class, but also between a native variation and a nonnative variation of the native language, i.e., American English from British English (Jusczyk, 2002; Nazzi, Jusczyk, & Johnson, 2000). These studies show how adept infants are towards identifying, distinguishing, and processing auditory stimuli in the form of spoken language.

Despite the rapid development of the aptitude to discern languages from one another, the evidence presented above does not establish an auditory dominance or even a preference. In order to argue the existence of such a dominance, infants not only have to be able to process both auditory-verbal and auditory-nonverbal stimuli, but the processing of both types of auditory stimuli should overshadow that of visual stimuli. In the case of auditory-nonverbal stimuli, an auditory dominance was only found in situations where the visual stimuli was static (Lewkowicz, 1985, 1988a, 1988b), whereas a visual preference was found when moving objects were used as visual stimuli (Spelke, 1979, 1981; Spelke, Smith Born, & Chu, 1983; Walker-Andrews, 1986). As mentioned above, Lewkowicz (1988a) conducted a series of experiments with 6-month-old infants where participants were presented with a flashing checkerboard as visual stimuli and a pulsating tonal sound as auditory stimuli. Infants were habituated to both visual and auditory stimuli simultaneously. During habituation, infants expressed an auditory dominance in several situations including when change rates varied and length of stimuli varied; meaning that the auditory processing blocked the processing of the visual stimuli. However, Lewkowicz (1988b) showed that 10-month-olds were also able to process aspects of the visual stimuli when the visual stimuli was presented in a more salient manner, i.e., when the visual stimulus was brighter.

Further investigations have shown evidence that auditory preference is still present in later childhood at four years of age (Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003). Sloutsky and Napolitano (2003) conducted a series of experiments where they presented infants with visual (either landscape pictures or geometric shapes) paired with auditory stimuli (tonal sounds) in a training phase. In the first pair of experiments, the goal was to consistently present a single visual stimulus with an auditory stimulus on one side of the screen with a distractor pair of stimuli played afterwards on the other side of the screen. In a test phase, Sloutsky and Napolitano (2003) tested their sensory attention by paring the trained visual stimulus with a novel auditory stimulus and in another condition with a novel visual stimulus paired with the trained auditory stimulus, then asking the children to select the equivalent pair. The results indicated that children rely more on the auditory stimuli for equivalence judgement tasks. The second set of experiments focused on how children encoded simultaneously presented audio and visual stimuli, focusing on both or on a single modality. In this set of experiments, Sloutsky and Napolitano (2003) found that 4-year-old children more readily encoded auditory stimuli than visual stimuli. Thus, both infants and children indicate an auditory dominance. More recent research suggests that the auditory dominance ends around 6 years of age with the transition to a visual preference beginning at school age (Nava & Pavani, 2013).

Now that evidence for an auditory dominance over visual processing has been established in infants and children, it is also important to discuss the underlying mechanisms that lead to this dominance. Robinson and Sloutsky (2019) identified two mechanisms that could contribute to the auditory dominance. The first mechanism is overshadowing. Overshadowing describes the difficulties that infants and young children have in attending to both auditory and visual stimuli when presented simultaneously (Robinson & Sloutsky, 2019). Given the auditory modality's importance early in development, infants and children might be more likely at attend to auditory information, or delay attention to visual features, leading to an auditory dominance over visual processing (Robinson & Sloutsky, 2004, 2007, 2010a; Sloutsky & Napolitano, 2003). The second mechanism underlying auditory dominance found by Robinson and Sloutsky (2019) is response competition. The response competition account refers to older children's ability to encode both auditory and visual stimuli in a cross-modal setting, but that a processing hierarchy for the specific processing of auditory and visual information may lead to an auditory dominance (Robinson & Sloutsky, 2019). Although the authors do not elaborate on this mechanism in great detail, a priority on attenuating to auditory information is provided in the example of children comparing two sets of audiovisual pairs. As such, children may first compare the two auditory stimuli before comparing the visual stimuli, leading to an auditory dominance (Robinson & Sloutsky, 2019). The overshadowing account and the response competition account are presented as possibilities for different developmental stages, with overshadowing being more pertinent to infants and response competition more relevant to older children (Robinson & Sloutsky, 2019).

In addition to auditory dominance, the abilities of infants to apply statistical learning to auditory stimuli is also important for their auditory processing. Seminal work by Saffran, Aslin, and Newport (1996) presented 8-month-old infants with trisyllabic nonsense words without word-boundary indicators other than transition probabilities, higher within words than between words, for two minutes. After familiarization, infants were tested on these newly learned syllable strings with matching strings (words) and strings consisting of the correct syllables, but in the incorrect order (nonwords). Listening times for the nonwords was significantly longer than those of words (Saffran et al., 1996). This novelty effect indicates that infants were able to identify the difference between the learned words and the novel nonwords after only a two-minute familiarization phase (for a similar discussion on the novelty effect, see Emberson et al., 2019). In addition to being able to apply statistical learning to auditory-verbal stimuli, Saffran, Johnson, Aslin, and Newport (1999) were able to show that 8-month-old infants can also apply statistical learning to auditory-nonverbal stimuli. The authors tested statistical learning with tonal sequences in a similar way to Saffran and colleagues (1996). Not only were infants able to distinguish learned sequences from novel sequences, but the authors speculate that auditory-verbal stimuli are not a requirement for auditory statistical learning in infancy (Saffran et al., 1999). The ability to apply statistical learning to auditory stimuli is not only found in older infants. Teinonen, Fellman, Näätänen, Alki, and Huotilainen (2009) provided evidence that neonates, tested in the first week after birth, were able to segment pseudowords out of a stream of syllables. In comparison to other modalities, it has also been shown that in unimodal statistical learning that 8- to 10-month-old infants are more successful in auditory statistical leaning than in visual statistical learning (Emberson et al., 2019). In a general glance, we can see that infants are not only substantial statistical learners in the auditory modality, when stimuli are presented sequentially, but also that they concentrate on auditory information in cross-modal situations to the point where auditory processing overshadows visual processing.

2.2.2 Visual Processing

As already stated above, Herschkowitz (1988) identified that a fetus in utero already reacts to visible light around 23 weeks gestation, which is very close to the fetal developmental track of audition. However, there is a poverty of stimulus input while the fetus is in utero in comparison to the auditory modality. This, in turn, leads to processing contrasts in neonates, whereas visual processing requires several months after birth in order to reach the same level of maturity as auditory processing (Banks & Salapatek, 1983). Despite this, neonates demonstrate preferential looking and selective attention, if only for short periods of time (for a review, see Reynolds, 2015). Over the course of the first year of life, visual processing abilities, including voluntary control and maintenance of visual attention, rapidly develop and improve (for reviews, see Colombo, 2001; Reynolds, 2015; Reynolds, Courage, & Richards, 2013).

With regard to the development of visual processing, two main abilities will be focused on in an attempt to compare visual processing to auditory processing – face recognition and object recognition. Unlike neonates' preference of the mother's voice, which a fetus is capable of perceiving in utero, a neonate has not seen the mother's face prior to birth. Despite this, Field, Cohen, Garcia, & Greenberg (1984) were able to indicate that neonates, at an average age of 45-hours-old, were able to discriminate between the face of the mother and a stranger. This, of course, was only done with habilitation to the mother's face or the mother's face and voice. Although the voice cues were unnecessary for the discrimination, the authors suggest that the distinction with such a limited experience may also stem from other multimodal cues such as odor (Field et al., 1984). Further research on facial recognition indicates that 1- to 3-month-old infants are capable of familiarizing new faces and recognizing them, but only 3-month-old infants were able to create average face prototypes and recognize them (de Haan, Johnson, Maurer, & Perrett, 2001). De Haan and Nelson (1997) provided evidence that 6-month-olds' neurocognitive processing of the mother's and face differed from that of the stranger's face depending on how similar the faces were. On the point of object recognition, infants have been found to be able to detect similarities in both spatial relations as well as features of novel visual objects by of 3 to 4 months of age (Bomba & Siqueland, 1983; Quinn, Slater, Brown, & Hayes, 2001). Furthermore 5-month-old infants have been shown to recognized familiarized novel objects in both color and black and white pictures as well as to recognize three-dimensional objects after familiarization of two-dimensional representations (DeLoache, Strauss, & Maynard, 1979). These studies give a brief overview of visual processing during the first few months of life.

Now that we have a basis for visual processing during infancy, another important aspect is infants' ability to apply statistical learning to visual stimuli. Seminal work in the field of visual statistical learning during infancy was conducted by Kirkham, Slemmer, and Johnson (2002) with 2-, 5-, and 8-month-olds. In a familiarization phase, infants were presented with pairs of colored shapes, six shapes in total, with similar transitional probabilities as in Saffran and colleagues (1996) and Saffran and colleagues (1999). After familiarization, infants were tested on matching pairs and novel sequences in a randomized order. All age groups, 2-, 5-, and 8-month-olds, displayed a novelty preference for the randomized sequences in the test phase, where there was no significant differences in the performance between any of the age groups (Kirkham et al., 2002). This suggests that infants as young as 2 months of age are able to reliably apply statistical learning to visual sequences and display novelty preferences for variations of presented stimuli. However, there is also evidence that even younger infants are capable of the same. Bulf, Johnson, and Valenza (2011) conducted a similar experiment with 1-to-3-day-old infants. Although the neonates reached the habituation criterion in the familiarization phase in both the high demand (six objects) and low demand (four objects) conditions, only the infants in the low demand condition displayed an above chance novelty preference for the randomized sequences in the test phase (Bulf et al., 2011). Despite only confirming statistical learning in neonates with shorter visual shape sequences, this finding does show that infants are able to apply visual statistical learning very early in life, as they are with the auditory modality (cf. Teinonen et al., 2009).

To understand the mechanisms behind infant visual statistical learning, it is important to note that Emberson and colleagues (2019), as stated above, identified separate mechanisms behind statistical learning, a novelty preference in the auditory modality and a familiarity preference in the visual modality. These findings in the visual modality, however, contrast those of Bulf and colleagues (2011) as well as Kirkham and colleagues (2002). The first factor that could have contributed to this is age. Raviv and Arnon (2018) tested, albeit older children, 5-to-12-year-olds in both auditory and visual statistical learning. They found that visual statistical learning improved over time, whereas the auditory statistical learning did not change (Raviv & Arnon, 2018). It could be that the older infants in the study by Emberson and colleagues (2019) have developed their visual statistical learning abilities to utilize a different mechanism than the younger infants in Bulf and colleagues (2011) and Kirkham and colleagues (2002). The other factor that could have contributed to the different mechanism is the stimuli used in the experiments. Whereas younger infants were familiarized with sequences of colored shapes (cf. Bulf et al., 2011; Kirkham et al., 2002), the older infants were presented with more complex visual stimuli in the form of pictures of human faces (cf. Emberson et al., 2019). The use of a more complex visual stimuli could have affected the preference of the infants and modulated their learning. Regardless of which preference the infants displayed, it is clear that infants are very capable of applying statistical learning to visual stimuli in the first year of life.

2.2.3 Multimodal Processing

Although children have a general preference for the auditory modality and cross-modal learning situations were already presented in Subchapter 2.2.1, the focus of this chapter is on the general aspects of multimodal processing in infancy and infants' abilities to apply statistical learning in multimodal situations. As already described above, infants are capable of multimodal processing in situations with static visual stimuli, when the rate of change is the same between visual and auditory stimuli, (Lewkowicz, 1988a, 1988b) and when the visual stimuli is a moving object (Spelke, 1979, 1981; Spelke et al., 1983; Walker-Andrews, 1986). Despite the fact that infants are able to process multimodal situations, there is evidence that when the information presented in the visual and auditory modalities are not coordinated processing costs can be incurred.

In multimodal learning situations, it has been found that familiar auditory stimuli, i.e., familiarized labels, slow down visual processing in 10-month-olds; however, unfamiliar auditory stimuli slow down visual processing even more (Robinson, Howard, & Sloutsky, 2005; Robinson & Sloutsky, 2007; Sloutsky & Robinson, 2008). Further research on 10-month-olds concluded that the cost of processing visual and auditory stimuli was disproportional in changing over time during multimodal presentations and that habituation criterion took longer to meet in multimodal stimulus presentations than in unimodal auditory presentation (Robinson & Sloutsky, 2010b). The authors attribute the differences in the changes of processing times to auditory dominance due to the fact that, in multimodal processing, pre-familiarizing infants to the auditory stimuli lead to reduced processing time of the auditory stimuli, whereas pre-familiarizing infants to visual stimuli did not reduce the processing time of visual stimuli (Robinson & Sloutsky, 2010b). These studies show that although infants are successful in applying statistical learning in cross-modal learning situations, the multimodal presentation of both visual and auditory information incurs processing costs.

2.3 Sensory processing in adults

Sensory processing continues to be important in adulthood. Auditory processing allows for verbal communication between two interlocutors, while visual processing allows for spatial orientation and navigation. As the development of both auditory and visual processing was presented in Subchapter 2.2, this Subchapter focuses on unimodal processing in adulthood, statistical learning within the auditory and visual modalities, short-term and working memory within the modalities, as well as these processes and effects in multimodal processing and learning situations.

2.3.1 Auditory Processing

As described at the beginning of this Chapter, adults tend to show a preference for visual information over auditory information in cross-modal situations (cf. Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003). This alone, however, does not mean to suggest that adults are poor auditory processors or learners. Saffran (1999) conducted two experiments with adult participants using auditory-verbal and auditory-nonverbal speech segmentation. In both experiments, adults were able to use the transitional properties of both auditory-verbal and auditory-nonverbal sequences in order to familiarize repeated sequences and identify novel randomized sequences (Saffran et al., 1999). This can be further supported by Conway and Christiansen (2005), where adult participants were able to apply statistical learning to an artificial grammar learning situation and, after a training phase, were able to correctly identify 75% of the test sequences over a control group, who only correctly identified 44% of test sequences. Saffran (2002), conducted a similar experiment and found that adults were able to apply statistical learning to auditory information using transitional properties. However, it was speculated that auditory statistical learning varies from visual statistical learning when it comes to predictive relationships between the items.

Saffran (2002) hypothesized that auditory statistical learning benefits from the sequential presentation of stimuli due to the fleeting nature of sounds, as they do not continue to exit over time, as for example linguistic information, musical patterns, sounds and tones produced across species, and other various environmental sounds (Saffran, 2002). Furthermore, this sequence of auditory stimulus presentation can also be seen in classical Pavlovian conditioning experiments (for a review, see Boakes & Costa, 2014). Classical conditioning considers two specific types of temporal presentations: delay conditioning and trace conditioning. Delay conditioning describes situations where the presented stimuli overlap in time (comparable to the simultaneous presentation in the current dissertation), whereas trace conditioning describes conditioning where there is a gap in time between the offset of the first stimulus and the onset of the next stimulus (comparable to the sequential presentation in this dissertation). Weike, Schupp, and Hamm (2007) provided empirical evidence which suggests that delay conditioning is a more implicit learning process and trace conditioning might require more explicit task knowledge. This attribute of the statistical learning of auditory information is vital to the remainder of the current dissertation. As such, not only the processing of auditory information is important, but also the short-term working memory and recall of auditory information associated with auditory processing are fundamental aspects necessary for the current work.

In terms of the memory effects of auditory processing, several aspects will be considered, including immediate recall and recollection, short-term working memory, and the effects of both as pertaining to auditory-verbal and auditory-nonverbal input (for a detailed review, see Snyder & Gregg, 2011). De Gelder and Vroomen (1997) conducted a series of experiments of serial presentation of four types of stimuli – auditory-verbal, auditory-nonverbal, written word, and a line-drawing. After presentation of stimuli, the authors found a large recency effect for spoken words (auditory-verbal) and an intermediate recency effect for environmental sounds (auditory-nonverbal), whereas no significant recency effect was reported for either of the visual conditions (De Gelder & Vroomen, 1997). However, this effect seems to decrease rapidly over time. Bigelow and Poremba (2014) provided evidence that auditory retention over time is weaker than visual and tactile recollection. Participants indicated significantly weaker auditory recollection on the same day as the familiarization and the day after the familiarization, but no significant differences between the modalities were reported for the week after familiarization (Bigelow & Poremba, 2014). As it pertains to more complex stimuli, Cohen, Horowitz, and Wolfe (2009) suggest that auditory short-term memory is inferior to visual short-term memory; however, the authors do not claim a generally impoverished long-memory in the auditory modality. These studies suggest that auditory short-term memory is weaker than visual short-term memory. However, in order to understand this further as it pertains to the current dissertation, it is important to also understand the mechanisms behind auditory short-term memory.

One of the most important factors for auditory short-term memory is retaining those auditory objects in working memory. Joseph, Kumar, Husain, and Griffiths (2015) identified that for auditory-nonverbal information, performance for memory dropped with each additional object added. Furthermore, auditory working memory seems to be based on objects and memory retrieval for individual features of auditory objects weakened auditory working memory, presuming that an object consists of multiple features (Joseph et al., 2015). There are two main mechanisms believed to contribute to the maintenance of auditory information, namely the phonological loop, a passive-maintenance system, and a cognitive process related to auditory imagery, an active-maintenance system (cf. Soemer & Saito, 2015). The phonological loop is described as a limited component of working memory that is utilized as a brief storage system to maintain information by means of vocal or subvocal rehearsal (for a review, see Baddeley, 2012). The phonological loop is said to have a few vital tools at its disposal, which provide support for maintaining auditory information including: the phonological similarity effect, the word length effect, articulatory suppression, irrelevant sounds effect, and retaining serial order (Baddeley, 2012). It is also important to note that Baddeley (2012) states that the phonological loop can also be used for visual stimuli in the form of written words, lip-reading, and sign language, as well as auditory-nonverbal information. A set of experiments conducted by Soemer and Saito (2015) investigated the claim for auditory-nonverbal information being maintained by an articulation-based rehearsal system, as timbre, or tonal quality, is not maintained in the phonological loop. Participants were presented with two-four artificial sounds differing in timbre. After either a short or a long delay, a single sound was played participants had to judge if the sound was part of the set while also conducting a pitch comparison and a brightness comparison task. The additional auditory imagery task, the pitch comparison, disrupted the maintenance of timbre, while the additional visual task, the brightness comparison, did not disrupt timbre maintenance (Soemer & Saito, 2015). As the additional auditory imagery task interferes with the active maintenance of timbre in auditory-nonverbal stimuli, Soemer and Saito (2015) suggest that a similar cognitive mechanism, which is related to auditory imagery, might be involved in the maintenance of auditory-nonverbal information. Thus, there seem to be two mechanisms that contribute to the maintenance of auditory information in short-term memory – the phonological loop, for information that can be articulated, and a cognitive process related to auditory imagery, for auditory-nonverbal information which cannot be articulated.

2.3.2 Visual Processing

Despite displaying a preference for visual information in cross-modal audio-visual situations (cf. Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003), a visual preference in adulthood does not necessarily implicate that visual processing is superior to auditory processing in a unimodal capacity. As afore mentioned, this can be seen in the immediate recall of series lists in both the visual and auditory modalities, whereas a recency effect is found for auditory information and not for visual information (De Gelder & Vroomen, 1997). This can be further seen in the direct comparison between auditory and visual statistical learning. Conway and Christiansen (2005) found that in both the training and testing phase of a sequential statistical learning paradigm, where participants in the experimental group of auditory statistical learning (96% correctly matched items in training and 75% correct in the test sequences) statistically outperformed those in the experimental condition of visual statistical learning (86% correctly matched items in training and 63% correct in the test sequences). Despite this, adults are effective visual statistical learners, as visual object memory is generally stronger than auditory object memory (Cohen et al., 2009; Yuval-Greenberg & Deouell, 2007, 2009). In an unsupervised statistical learning experiment involving complex visual scenes, Fiser and Aslin (2001) were able to show that participants successfully acquired knowledge of absolute shape-position relations within an array, pairs of shapes independent of position, and the conditional dependencies of these shape co-occurrences in a rapid and automatic manner. In direct comparison with auditory statistical learning, Fiser and Aslin (2002) recreated the Saffran and colleagues (1996) with visual stimuli with adult participants. The authors found that adult participants are highly sensitive to the temporal structure of visually presented image sequences (Fiser & Aslin, 2002).

The experiments form Fiser and Aslin (2001, 2002) on the learning of visual object sequences can be categorized as visual artificial grammar learning in the same manner as the experiments of Saffran and colleagues (1996, 1999). This can be further compared to Saffran (2002), which tested the abilities of adults to apply statistical learning to both visual-verbal and visual-nonverbal stimuli in an artificial grammar learning situation. In these experiments, adult statistical learning was not influenced by the presence of linguistic stimuli, in the form of written pseudowords; however, it became apparent that, unlike in the auditory modality, visual artificial grammar learning was not affected by the predictive dependencies when visual artificial grammars were presented sequentially (Saffran, 2002). These findings are further supported by the findings of Aguilar and Plante (2014), who tested visual artificial grammar learning with adults with normal language and a language learning disability. The authors report that both test groups exhibited successful learning, but the language learning disability group indicated a nonsignificant trend of lower acceptance of correct items and a trend of slightly higher acceptance of incorrect items (Aguilar & Plante, 2014). The combined results of the studies presented in this section show that adults are successful visual learners; however, they are believed to be more efficient learners when visual information is presented simultaneously.

2.3.3 Multimodal Processing

Now that we have investigated unimodal processing and statistical learning in both the auditory modality and the visual modality, this Subchapter investigates multimodal audio-visual processing and statistical learning situations and the effect of multimodal processing on memory. Multimodal processing is more than the conglomeration of visual and auditory processing alone. Although unimodal information is vital, multimodal information from the environment may be more informative than unimodal information (Oviatt & Cohen, 2000; Sarter, 2006). In the beginning of this Chapter, two constraints were presented as they pertain to statistical learning – the developmental constraint and the stimulus constraint (Thiessen, 2010). The focus in this subsection is the stimulus constraint, which states that adults process auditory-verbal and auditory-nonverbal information differently in multimodal situations (cf. Thiessen, 2010). Thiessen (2010) concluded this from two experiments with adults on word learning – one by presenting visual information with spoken pseudowords and the other by presenting tonal sequences with the same visual input as the auditory-verbal condition. Both experiments presented the audio-visual situations in a simultaneous manner. The participants in the auditory-verbal experiment where auditory stimuli were paired with regular video preformed statistically better than in the irregular video or no video conditions (Thiessen, 2010). Furthermore, participants in the auditory-nonverbal experiment scored above chance in all three conditions, whereas there were no significant differences between them (Thiessen, 2010). Thus, adult participants were able to extrapolate auditory regularities; yet only benefited from cross-modal presentation when the auditory stimuli were linguistic (Thiessen, 2010). The benefits of multimodal learning can further be seen when comparing unimodal auditory, visual, and multimodal guided associative learning tasks (Eördegh et al., 2019). The authors were able to provide evidence that long-term memory for multimodal audio-visual stimuli is stronger than unimodal presentations; however, the effect is larger for matching audio-visual information than for mismatched combinations (Eördegh et al., 2019). Furthermore, memory performance was not affected with various realizations of temporal congruency; in other words, having the audio-visual information presented sequentially did not affect memory as compared to the simultaneous presentation (Eördegh et al., 2019). We can not only see that adults are capable of multimodal processing and learning, but that multimodal learning is quite flexible and memory seems to be less affected by changes in temporal congruity of stimuli.

Multimodal contexts significantly influence memory as compared to unimodal contexts; an increase in memory performance can be seen when the initial presentation of multimodal stimuli are semantically congruent, whereas a decrease in memory performance is seen when the initial presentation of multimodal stimuli are semantically incongruent (Matusz, Wallace, & Murray, 2017). This can be further supported by evidence in the brain itself. By looking at activation sources in memory tasks, multimodal perceptual traces (audio-visual) are distinctly different from unimodal perceptual traces (Matusz et al., 2017). Furthermore, research suggests that enhanced activation via multimodal (audio-visual) semantically congruent stimulation is reflected by greater activation of the respective cortical region than both unimodal stimulation and incongruent stimulation (Doehrmann & Naumer, 2008). Heikkilä, Alho, Hyvönen, and Tiippana (2015) provide further evidence for a facilitated memory effect when semantically congruent information in two modalities are encoded.

These studies show that the manner in which multimodal stimuli are encoded affects the retrieval of those memories. As such, the neural pathways and cortices yield greater activation in respective areas when information is being retrieved as compared to the retrieval of unimodal contexts. In sum, adults are able to effectively encode multimodal information despite its complexity as compared to unimodal information. Adults are not only more effective learners when multimodal information is presented in a semantically congruent manner, but memory performance is also enhanced for these semantically congruent contexts.

3 Word learning

Being able to communicate with others in any given language requires one to first learn that language. A large part of learning a language is acquiring lexical features of the language and mapping those onto objects and actions, be they concrete or abstract. Although most of language learning occurs during development, matured speakers also learn new words over the course of adulthood. In this Chapter, the phenomenon of word learning is presented as well as mechanisms behind word learning. The relation between sensory modality and language, as it pertains to word learning, are also explored. Finally, this chapter will also present the neurophysiological methodology of measuring word learning.

3.1 The phenomenon of word learning

In order to learn the meaning of a word, the learner must first understand the relationship between an object or an action and its spoken label. This may sound like a trivial task; however, it can take multiple exposures to the word in context in order to fully integrate the semantic connection into long-term memory. Understanding what exactly is being labeled is also not always a straightforward matter. This can be seen in the example of Quine (1960) where a guide in a foreign country labels a rabbit running across the road as *gavagai*. Although the first instinct is to think *gavagai* refers to 'rabbit', it could also very well refer to 'animal' or 'white' or something even more abstract. This example shows the difficulty that adults may face with learning new words, even with vast experience with language. However, children also face these difficulties, all the while making it look easy. Although children typically utter their first words shortly before their first birthday, between 10 to 12 months of age (cf. Johnson, 2016; Samuelson & McMurray, 2017), they begin learning these connections much earlier (e.g., Bergelson & Aslin, 2017; Bergelson & Swingley, 2012; Friedrich & Friederici, 2011, 2017; Tincoff & Jusczyk, 1999, 2012).

A seminal study on infant word learning showed that infants as young as six months of age link highly frequent sound patterns, such as *mommy* and *daddy* with their meanings (Tincoff & Jusczyk, 1999). In further experiments, Tincoff and Jusczyk (2012) investigated whether the early comprehension system of six-month-olds is also capable of supporting more generalized categorical associations. Utilizing a preferential looking task, the authors showed that six-month-olds comprehend the object-word pairs for *hands* and *feet* (Tincoff & Jusczyk, 2012). In an eye-tracking experiment with six-month-olds, Bergelson and Swingley (2012) found similar results on early word-meaning links pertaining to body parts and food items. However, the authors also noted on an apparent lack of improvement in looking time between 6- and 13-month-olds, only finding a much larger proportion of target looking time in 14-month-olds. These studies show that early word learning, or at least associating a word with an object, is facilitated by highly frequent occurrences of the respective objects in everyday life. Infants explore their surroundings and, in doing so, experience rich sensory input which helps build perceptually based categories capable of supporting long-term retention of semantic meaning (Maouene, Hidaka, & Smith, 2008; Smith, Maouene, & Hidaka, 2007).

A main focus of research on early word learning relates to the saliency of the object in the infant's focus of attention (Clerkin, Hart, Rehg, Yu, & Smith, 2017; Samuelson & Smith, 1998). Human attention is, however, flexible and can attend to various perceptual properties of stimuli (Jones & Smith, 2002; Jones, Smith, & Landau, 1991; Nosofsky, 1986; Smith, Jones, & Landau, 1996). However, the only modality of perceptual salience that has been investigated in early word learning is the visual modality. Although we know that rich sensory input helps in building categories for word learning, there is still a question whether the modality in which the object is presented not only has an impact on how it is perceived, but whether it also has an effect on its saliency and learning. Research to date leaves the role of object modality in word learning greatly unexplored. Many word learning models and theories have embraced perceptual saliency on the stimulus side and attention on the side of cognitive function as integral factors. Perceptual saliency is thought to have an impact on word learning via its role in guiding the attentional system. A prime example of this is the Emergentist Coalition Model (ECM). According to the ECM, children are sensitive to multiple cues in early word learning – attentional, social, and linguistic (Hollich, Hirsh-Pasek, & Golinkoff, 2000). Attentional cues are considered to be among the earliest influences on word learning. Factors in this category are domain-general and include perceptual salience, temporal congruity, and novelty. One of the main hypotheses of the ECM is that young children are sensitive to attentional cues much earlier than they become sensitive to social and linguistic cues, which occurs around 12 months of age (Hollich et al., 2000). Smith, Jones, and Landau (1996) presented the *dumb attention hypothesis* which states that attention in children's early word learning is driven by perceptual salience of object properties alone and information about the function of the object does not influence the naming of the object at all. Results confirmed that 3-year-old children's attention in the novel naming task was solely influenced by the perceptual salience of the object, while information about the function of the object was ignored (Smith et al., 1996). Later studies by Smith and colleagues provided evidence that the visual input from an egocentric view plays a major role for learning, further supporting the explanatory power of perceptual salience (Smith, Jayaraman, Clerkin, & Yu, 2018; Yu & Smith, 2012). By exploring paradigms presenting objects in other than the visual modality, e.g., audition, we can begin to understand how different types of perceptual information affect word learning.

3.1.1 Word learning and learning mechanisms

We have now defined that humans, especially infants, are very adept at acquiring new words, despite the difficulty of the task. In this Subchapter, the focus is not on why humans can learn words, but rather on how words are acquired. As mentioned above, there are several different types of cues that infants and children attend to, which aid in the acquisition of meaning and word learning – attentional, social, and linguistic cues (cf. Hollich et al., 2000). In looking at the mechanisms behind word learning in infants, these cues, although sometimes more narrowly focused, can be found in many different theories. Attentional cues can also be found in mechanisms such as shape bias, where words for solid shapes is extended to similar novel shapes (Colunga & Smith, 2005). These attentional cues can also be narrowed to concepts such as saliency, which comprise both attention and perception together to highlight how infants perceive and attend to objects based on brightness, color, motion, size, illumination, (e.g., Pruden et al., 2006; Smith et al., 1996; Werker et al., 1998) and even novelty (Samuelson & Smith, 1998; for a review on saliency in word learning, see Wildt, Rohlfing, & Scharlau, 2019). Linguistic cues are also an important focus in research and theories behind the mechanisms of word learning, many with a focus on bootstrapping. Bootstrapping assumptions focus on how language is acquired with the aid of input features (for a review, see Höhle, 2009). There are many different types of bootstrapping including: distributional bootstrapping, where nonprosodic segments are statistically analyzed to aid in the parsing of input (e.g., Gerken, 1996; Höhle, Weissenborn, Kiefer, Schulz, & Schmitz, 2004; Mintz, Newport, & Bever, 2002); semantic bootstrapping, in which semantic-syntactic associations aid in linguistic categories (e.g., Pinker, 1984, 1987); syntactic bootstrapping, in which context is used to aid in understanding meaning (e.g., Gleitman, 1990; Naigles, 1990; Waxman & Booth, 2001); typological bootstrapping, where the focus is on crosslinguistic variation in form-function relations (e.g., Bowerman & Choi, 2001; Slobin, 2001); and prosodic bootstrapping, where prosodic cues aid in the acquisition of underlying grammatical organization (for a detailed review, see Höhle, 2009). And finally, the social cues typically are paired with attentional cues, for example in classical parent-child interactions with learning from a form of scaffolding supported by others (e.g., Bruner, 1983), or with linguistic cues such as social-pragmatic theories, which focus on the communicative intentions of others to learn words without linguistic constraints (e.g., Tomasello, 2000).

However, these theories, as aforementioned, focus on how infants and children acquire words and create abstractions such as categories. Within the scope of this dissertation, this type of long-term learning is not necessary, although may be applicable. As such, there are two categories of mechanisms that combine various aspects of many of these mechanisms. These two approaches focus on multiple encounters with a stimulus and its label. With each additional experience, a more defined relationship between the object and its label is created. The two approaches that address word learning in this capacity are inference-based models and theories and associative-based models and theories.

Inferenced-based models and theories of word learning are also known as hypothesis testing or hypothesis elimination (e.g., Bloom, 2000; Pinker, 1984, 1989; Siskind, 1996; Xu & Tenenbaum, 2007). In these types of models and theories, it is assumed that a word maps onto a single concept and the learner creates a hypothesis as to what the word is referring to, even with a single presentation of a novel word and a novel stimulus (cf. Xu & Tenenbaum, 2007). An early example of inference-based word learning can also be seen in the mapping problem as described by Quine (1960). The *gavagai* problem is described as a learner creating a hypothesis on what a new word could mean and then subsequently testing it in further instances and encounters with the stimulus (Quine, 1960). For hypothesis elimination models and theories, the hypotheses that are not confirmed are then eliminated as being further possibilities. The concept of hypothesis elimination is a classical idea, which has also been adapted to more modern concepts of inference-based learning. Xu and Tenenbaum (2007) presented an revised hypothesis testing model of word learning by combining hypothesis elimination with a Bayesian inference framework. Within this model, hypotheses are not eliminated as they are tested, but rather the probability of correctness is adjusted across observations (Xu & Tenenbaum, 2007). Both hypothesis elimination and the Bayesian inference models of word learning have been tested in both adults and children using the concept of taxonomic coherence, in which a generalization is created across a specific taxonomic concept with items either belonging to the group or not belonging to the group. While both adults and 3-to-4-year-old children were successful in word learning via Bayesian inference (Xu & Tenenbaum, 2007), only adults were able to successfully extrapolate meaning via hypothesis elimination (Sloutsky et al., 2017).

As mentioned above, word learning can also be achieved through a different mechanism other than inference-based, namely through associative learning. Association-based models and theories of word learning have been used to describe how infants and children are able to acquire novel word meanings (e.g., Colunga & Smith, 2005; Elman, 2009; Mayor & Plunkett, 2010; McMurray et al., 2012; Regier, 2005; Samuelson & Smith, 1999; Sloutsky et al., 2017; Yu & Smith, 2007, 2012). Unlike the inference-based models and theories of word learning, associative-based models and theories build up the connections between words and their referents through repeated exposure. As aforementioned, these types of word learning accounts are related to statistical learning (see Chapter 1 and Chapter 2). Although associative word learning and inference-based word learning are fundamentally different, some inference-based accounts do argue that associative learning could be a possible mechanism in early word learning (e.g., Golinkoff & Hirsh-Pasek, 2006; Namy, 2012; Nazzi & Bertoncini, 2003), providing evidence that these accounts offer various explanations for word learning and many not be completely mutually exclusive. Other accounts provide evidence that adults, unlike children, are capable of utilizing both inference-based and associative-based word learning in order to acquire new words (cf. Sloutsky et al., 2017). For this dissertation, the association-based account of word learning is enough for short-term retention of novel object-word pairs. Association-based accounts for word learning have also been shown to be similar in both infants and adults, which further makes it a plausible mechanism to test in order to have a more well-defined understanding of how words in infancy and adulthood are learned under similar conditions.

3.1.2 Word learning and sensory modality

As previously mentioned, we use language to communicate our experiences with others. This means, of course, that language has the capacity in which to convey these sensory observations between individuals. Although there are commonalities between languages, for example the existence of words and the use of some form of syntax, how languages label concepts, both concrete and abstract, differ. Thus, the manner in which sensory observations are communicated may also be realized in differently between languages. Over the past few years, it has been found that there is a hierarchy of perceptual modality norms which can be found in the semantics of a given language. Perceptual modality norms have been found for individual languages, including English (Lynott et al., 2020; Majid et al., 2018; Winter et al., 2018), Dutch (Speed & Majid, 2017), Italian (Vergallito et al., 2020), French (Chedid et al., 2019), Russian (Miklashevsky, 2018), and Mandarin Chinese (I.-H. Chen et al., 2019). The perceptual modality norms for Indo-European languages as well as Mandarin Chinese indicate that more words are allocated to the visual modality than to any other modality. Pertaining to the other modalities, there is no evidence of any hierarchy across languages. However, the evidence provided from these studies do not mean to paint a broader picture that is applicable to all languages or even all Indo-European languages. Majid and colleagues (2018) tested several languages across the globe, including indigenous languages in many countries as well as multiple sign languages, by means of the codability of perceptual modalities. Codability, in this sense, is the ability of a speaker to convey what is being perceived to another speaker based on a set of examples. Majid and colleagues (2018) were able to identify that humans are not universally able to convey visual and auditory experiences better than other modalities across languages, as some languages had higher mean codability indexes in the gustatory modality than both the auditory or visual modalities. Majid and colleagues (2018) also identified that in Farsi, an Indo-European language, also did not replicate the dominance of the visual modality over the other modalities as was the case in other Indo-European languages. Although the perceptual modality norms have not been mapped for all languages nor have they been mapped for the entirety of a single language, these findings do indicate that humans are able to convey their experiences in several modalities effectively. For the scope of this dissertation, perceptual modality norms, within a given language, help shed light on how words are allocated across modalities. As sensory processing and language are intertwined, there may be an overlap of a learner of a particular language and the sensory processing abilities of humans in infancy and adulthood. Thus, the field of sensory linguistics can help shed light not only on the sensory aspects of a language, but also how a learner may go about learning the language.

3.2 Methodological approach to measuring word learning

Within the scope of this dissertation, word learning is measured by the electrophysiological ERP component known as the N400. Utilizing the ERP technique allows the neural processing changes over time to be recorded and measured time-locked to the onset of a particular stimulus, in this case the spoken word. The N400 component is a particular negative waveform that typically peaks around 400 ms after the onset of a stimulus (for a review, see Kutas & Federmeier, 2011). Kutas and Hillyard (1980) originally discovered and defined the N400 as a more negative ERP response to a stimulus which is presented in a contextually violated condition in comparison to an ERP response to a control stimulus presented as expected within the context. Although the N400 was first used to identify semantic violations within sentence context, this component has also been used to measure semantic priming including word learning in adults (e.g., Angwin, Phua, & Copland, 2014; Bakker, Takashima, van Hell, Janzen, & McQueen, 2015; Balass, Nelson, & Perfetti, 2010; Bermúdez-Margaretto, Beltrán, Cuetos, & Domínguez, 2018; Borovsky, Elman, & Kutas, 2012; S. Chen, Wang, & Yang, 2014; François, Cunillera, Garcia, Laine, & Rodriguez-Fornells, 2017; Kaczer et al., 2018; Key, Molfese, & Ratajczak, 2006; Lau, Holcomb, & Kuperberg, 2013; Mestres-Misse, Rodriguez-Fornells, & Munte, 2007; Nobre & McCarthy, 1994, 1995; Shtyrov, 2011; van Petten & Rheinfelder, 1995) and children (e.g., Abel, Schneider, & Maguire, 2018; Borgström et al., 2015b; Borgström, Torkildsen, & Lindgren, 2015a; Friedrich & Friederici, 2008, 2011, 2017; Friedrich, Wilhelm, Born, & Friederici, 2015; Friedrich, Wilhelm, Mölle, Born, & Friederici, 2017a; Junge et al., 2012; Mills et al., 2004; Mills, Plunkett, Prat, & Schafer, 2005; Torkildsen et al., 2009, 2008; Torkildsen, Syversen, Simonsen, Moen, & Lindgren, 2007b). Not only can the N400 be used to test contextually violated real words, but the component can also be used to measure legal nonwords (pseudowords) during learning and in contextually violated tests after learning, but the component is not as sensitive to the use of illegal nonwords (Nobre & McCarthy, 1994). Mestres-Misse, Rodriguez-Fornells, and Munte (2007) found that the N400 can also be used to measure the acquisition of novel words over time, but also that even after three exposures, the N400 amplitudes were indistinguishable from those of real words in semantically violated contexts, even when words were presented in isolation. Although the N400 component is found in both adults and children, there are differences between the two realizations of the component.

In infants, ERPs typically show that 6- to 9-months-olds are sensitive to novel audio-visual object-word associations (e.g., Friedrich & Friederici, 2008, 2011; Junge, Cutler, & Hagoort, 2012; Parise & Csibra, 2012); yet, there is also evidence that 3-month-old infants are capable of building associative relationships between a visual novel object and a novel spoken pseudoword, but not retaining these associations over a 24-hour time period as 6- and 14-month-old infants do (Friedrich & Friederici, 2017). However, unlike with the N400 component in adults, several different ERP components have been found which indicate, potentially at different representation levels, word-form familiarity effects (during learning) or violated expectation effects (after learning) in word learning studies: the N200-500 (Friedrich & Friederici, 2008; Kooijman, Hagoort, & Cutler, 2005; Mills, Coffey-Corina, & Neville, 1993; Thierry, Vihman, & Roberts, 2003; Torkildsen et al., 2009) and the N400-like component (Borgström et al., 2015b; Friedrich & Friederici, 2008; Junge et al., 2012; Torkildsen et al., 2008, 2007a), respectively (for a more detailed discussion, see Cosper, Männel, & Mueller, 2020). Word-form familiarization in infants is typically seen slightly earlier than the N400 and with a different topological distribution. The word-form familiarity effect has been typically reported around 200-500 ms after the onset of the word and with a fronto-central distribution, measured by a modulation in the amplitude in a negative direction with repeated word presentation or word-form knowledge (for a review, see Teixidó, François, Bosch, & Männel, 2018). Previous studies have additionally reported on later occurring word-form repetition effects, which in the case of Torkildsen and colleagues (2009) was reported from 400-1400 ms after the onset of the word. The N400 component has also been used as a measure for word learning in infant studies. As its timing and topography do not always correspond to the properties of the adult N400, we term it N400-like component in the context of infant studies. The N400-like component in association-learning paradigms, similar to the adult ERP component, reflects a violation of expectation and occurs with more negative amplitude waveforms for violated than matching conditions. During mapping in such paradigms, object-word associations are being formed by means of repetition. As such, repeated exposure to the associated pairs decreases the amplitude of the N400 over time. Thus, the N400 violated expectation effect does not necessarily reflect semantic integration, but can also be explained by a simple spreading activation. It is, however, important to note that the infant N400-like component, in comparison to the adult N400 component, is slightly temporally delayed (Junge et al., 2012). Infant N400 effects have been found in various time windows, with reports beginning at 300 ms (Friedrich et al., 2015) and even stretching to 1200 ms (Torkildsen et al., 2008).

4 Current set of studies: Hypotheses within and between age groups

In order to investigate auditory associative word learning, several EEG studies were conducted within the scope of this dissertation. All experiments were based on Friedrich and Friederici (2008, 2011), an associative word learning paradigm which presented infants with pictures of colorful novel objects and spoken pseudowords. As the papers within this dissertation were the first to assess auditory associative word learning, retention over a 24-hour period was not tested, but rather short-term recall was tested in these cases in order to assess the capabilities of learners across the lifespan. The infant and the adult experiments will briefly be presented with the hypotheses pertaining to the respective age groups; however, more detailed information can be found in Part II of this dissertation. Subsequently, the hypotheses between the age groups will also be presented. The goal of this cumulative dissertation is to assess the auditory associative word learning abilities between the age groups as well as how aspects of word learning and the modulation of the N400 are also affected by aspects of the experimental setup.

Chapter 5 presents a study conducted by Cosper, Männel, and Mueller (2020) on auditory associative word learning in 10- to-12-month-old infants. Here, infants were presented with real environmental sounds in the form of short sound clips taken from the NESSTI database for environmental sounds (Hocking, Dzafic, Kazovsky, & Copland, 2013) and spoken disyllabic pseudowords in either a consistently or inconsistently paired manner. The sound-pseudoword pairs were presented in a temporally sequential manner, with an inter-stimulus pause of 600 ms, as it has been previously reported that statistical learning in the auditory modality benefits from sequential presentation (cf. Saffran, 2002). Immediately after an initial training phase, the participants were subjected to a testing phase where the consistently paired sound-pseudoword pairs were presented in a matching or violated manner, in order to determine if mapping was successful. We hypothesized that infants would be successful in mapping pseudowords onto auditory objects as indicated by an *N400-like effect*. Furthermore, we predicted that learning across the training phase would also be measured as indicated by a *word-form familiarity effect*, indicating that all pseudowords are familiarized during the training phase, and a *pairing consistency effect*, indicating that the consistent sound-word pairs would become more familiar over time as indicated by a reduction in the ERP amplitude.

Chapter 6 presents a similar study in adults. In this paper, Cosper, Männel, and Mueller (submitted) conducted a more systematic investigation of auditory associative word learning in direct comparison to audio-visual multimodal associative word learning. In this 2x2 design, four separate experiments were conducted by modulating the modality in which the novel object was presented (auditory vs. visual) and by modulating the temporal presentation of the sound-word pairs (sequential presentation vs. simultaneous presentation). The spoken disyllabic pseudowords and the sounds used in the auditory experiments were the same as in Cosper and colleagues (2020), whereas the objects used in the visual experiments were taken from the Amsterdam Library of Object Images (Geusebroek, Burghouts, & Smeulders, 2005) and the Novel Object and Unusual Name database (Horst & Hout, 2016). The sequential temporal presentation conditions presented the stimuli with an inter-stimulus pause of 600 ms, while the simultaneous temporal condition presented the object-pseudoword pairs with a 500 ms overlap of the object and the spoken pseudoword. In these experiments, two main hypotheses were presented. Hypotheses 1 states that the visual modality will elicit stronger effects of word learning as compared to the auditory modality (cf. Bigelow & Poremba, 2014; Cohen et al., 2009; Robinson & Sloutsky, 2004). Hypothesis 2 states that the visual modality will benefit from simultaneous presentation of stimuli, while the auditory modality will benefit from sequential presentation (cf. Conway & Christiansen, 2005; Saffran, 2002).

As aforementioned, the goal of this dissertation is to understand the change in auditory associative word learning across the lifespan. In order to do this, the findings from Chapter 5 and Chapter 6 are compared and contrasted. In doing so, there are three main hypotheses that guide the comparison. First, is the prediction that the modality in which the object is presented affects word learning differently in infants and adults, with infants being more capable to map novel pseudowords onto auditory objects than adults (cf. Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003; Thiessen, 2010). Second, the N400 component is modulated by the modality in which the object is presented, with modulation affecting infants and adults differently, in relation to the first hypotheses (cf. Junge et al., 2012; van Petten & Rheinfelder, 1995). Finally, in addition to modality, the temporal congruency in which objects and pseudowords are presented also affects word learning (cf. Boakes & Costa, 2014; Conway & Christiansen, 2005; Saffran, 2002; Weike et al., 2007).

Part II: Experiments

5 Paper I: In the absence of visual input: Electrophysiological evidence of infants' mapping of labels onto auditory objects

Abstract:

Despite the prominence of non-visual semantic features for some words (e.g., siren or thunder), little is known about when and how the meanings of those words that refer to auditory objects can be acquired in early infancy. With associative learning being an important mechanism of word learning, we ask the question whether associations between sounds and words lead to similar learning effects as associations between visual objects and words. In an event-related potential (ERP) study, 10- to 12-month-old infants were presented with pairs of environmental sounds and pseudowords in either a consistent (where sound-word mapping can occur) or inconsistent manner. Subsequently, the infants were presented with sound-pseudoword combinations either matching or violating the consistent pairs from the training phase. In the training phase, we observed word-form familiarity effects and pairing consistency effects for ERPs time-locked to the onset of the word. The test phase revealed N400-like effects for violated pairs as compared to matching pairs. These results indicate that associative word learning is also possible for auditory objects before infants' first birthday. The specific temporal occurrence of the N400-like effect and topological distribution of the ERPs suggests that the object's modality has an impact on how novel words are processed.

Authors: Samuel H. Cosper[1], Claudia Männel[2,3], & Jutta L. Mueller[1]

Affiliations:

[1] Institute of Cognitive Science, University of Osnabrück

[2] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig

[3] Department of Audiology and Phoniatrics, Charité–Universitätsmedizin Berlin

[4] Department of Linguistics, University of Vienna, Austria

Cosper, S. H., Männel, C., & Mueller, J. L. (2020). In the absence of visual input: Electrophysiological evidence of infants' mapping of labels onto auditory objects. *Developmental Cognitive Neuroscience*, *45*, 100821. https://doi.org/10.1016/j.dcn.2020.100821

This chapter has been redacted due to copyright issues; however, it can be accessed online.

Cosper, S. H., Männel, C., & Mueller, J. L. (2020). In the absence of visual input: Electrophysiological evidence of infants' mapping of labels onto auditory objects. *Developmental Cognitive Neuroscience*, *45*, 100821. https://doi.org/10.1016/j.dcn.2020.100821

6 Paper II: Mechanisms of word learning: Benefits from the visual modality and synchronicity of labeled objects

Abstract:

Modality is known to influence memory and learning, its specific role for language learning remains unclear. We investigated how word learning in adults is affected by (visual vs. auditory) object modality and synchronicity of stimulus presentation. In training phases, participants were presented either visual objects (real-world images) or auditory objects (environmental sounds) in temporal synchrony with or followed by novel pseudowords. Objects and pseudowords were paired either in a consistent or an inconsistent manner. In subsequent testing phases, the consistent pairs were presented in matching or violated pairings to test whether consistent object-pseudoword pairs were successfully associated with one another. While the visual modality yielded behavioral and ERP evidence for short-term retention of the novel words' associates in both the simultaneous and sequential conditions, similar results were found for the auditory modality only for the simultaneous condition. Thus, we argue for a visual and a synchronicity advantage for word learning.

Authors: Samuel H. Cosper[1], Claudia Männel[2,3], & Jutta L. Mueller[1,4]

Affiliations:

[1] Institute of Cognitive Science, University of Osnabrück

[2] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig

[3] Department of Audiology and Phoniatrics, Charité–Universitätsmedizin Berlin

[4] Department of Linguistics, University of Vienna, Austria

Paper submission status:

The manuscript has been submitted to *Psychonomic Bulletin & Review*, on 30.07.2020, and is currently under review. This unreviewed manuscript is published with the journals permission.

Mechanisms of word learning: Benefits from the visual modality and synchronicity of labeled objects

Authors: Samuel H. Cosper[1]*, Claudia Männel[2,3], & Jutta L. Mueller[1,4]

[1] Institute of Cognitive Science, University of Osnabrück

[2] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig

[3] Department of Audiology and Phoniatrics, Charité–Universitätsmedizin Berlin

[4] Department of Linguistics, University of Vienna, Austria

* Corresponding author at: Institute of Cognitive Science, University of Osnabrück, Wachsbleiche 27, 49090 Osnabrück, Germany; *email address:* samuel.cosper@uni-osnabrueck.de

Abstract:

Modality is known to influence memory and learning, its specific role for language learning remains unclear. We investigated how word learning in adults is affected by (visual vs. auditory) object modality and synchronicity of stimulus presentation. In training phases, participants were presented either visual objects (real-world images) or auditory objects (environmental sounds) in temporal synchrony with or followed by novel pseudowords. Objects and pseudowords were paired either in a consistent or an inconsistent manner. In subsequent testing phases, the consistent pairs were presented in matching or violated pairings to test whether consistent object-pseudoword pairs were successfully associated with one another. While the visual modality yielded behavioral and ERP evidence for short-term retention of the novel words' associates in both the simultaneous and sequential conditions, similar results were found for the auditory modality only for the simultaneous condition. Thus, we argue for a visual and a synchronicity advantage for word learning.

1 Introduction

We experience our world in a variety of sensory inputs including vision and audition. Language provides us with the means to express these experiences to each other, with words for visual, auditory, olfactory, haptic, and gustatory objects and events. However, research on the acquisition of words for concrete objects focuses on labeling visual objects, leaving many of the other modalities largely untouched (e.g., Friedrich & Friederici, 2008, 2011; Horst & Samuelson, 2008; Junge, Cutler, & Hagoort, 2012; Smith & Yu, 2008; Taxitari, Twomey, Westermann, & Mani, 2019; Werker, Cohen, Lloyd, Casasola, & Stager, 1998). This is surprising as modality is known to have an impact on the parameters and the outcome of learning and memory. Cohen, Horowitz, and Wolfe (2009) demonstrated that visual recognition is significantly better than auditory recognition. Similarly, Bigelow and Poremba (2014) provided evidence that short-term memory as well as delayed recognition were inferior in the auditory modality compared to both visual and tactile modalities. Within the visual and auditory modalities, processing differences have also been discovered pertaining to the temporal presentation of stimuli. Saffran (2002) showed that statistical learning (SL) in children benefitted from differing temporal congruencies between visual and auditory learning. Whereas visual SL benefitted from simultaneous presentation, auditory SL benefitted from sequential presentation (Saffran, 2002). In the current study we aimed to systematically investigate how an object's modality and its temporal occurrence in relation to its label affect word learning.

In order to learn the meaning of a word, the learner must first realize the relationship between an object and its specific label. In studies on the acquisition of labels for concrete objects, this relationship is often established by the simultaneous presentation of both object and word (e.g., Breitenstein et al., 2005; Friedrich & Friederici, 2008, 2011). Word learning, in such paradigms, is conceptualized as an associative process that links object and word by statistical co-occurrence (e.g., McMurray, Horst, & Samuelson, 2012; Sloutsky, Yim, Yao, & Dennis, 2017; Smith & Yu, 2008). Although this type of learning is also supplemented by other mechanisms (cf. Xu & Tenenbaum, 2007), we take associative word learning as a starting point for testing modality effects because of its fundamental importance across the lifespan (cf. Sloutsky et al., 2017).

The acquisition of new word meanings can be tested experimentally via electroencephalography (EEG) using the event-related potential (ERP) technique, specifically the N400 component. Kutas and Hillyard (1980) first described the N400 as a negative-going waveform around 400 ms after the onset of a target word which did not fit in its semantic context. Since then, the N400 has been employed in many studies on the processing of word form and meaning (see Kutas & Federmeier, 2011, for a review). The N400 component can be elicited both during and after learning; lending itself to test word-object associations during learning (Mestres-Misse et al., 2007) and violations of lexical-semantic expectations after learning (Friedrich & Friederici, 2008).

The current series of experiments explores how the mechanisms behind word learning are affected by object modality and the temporal congruency of object and label. In a 2x2 design, adapted from Friedrich and Friederici (2008, 2011) and Cosper, Männel, and Mueller (2020), four studies systematically test objects in the visual and auditory modality paired with spoken pseudowords in sequential and simultaneous manners (see Figure 1a). All experiments recorded both ERP responses time-locked to the onset of the pseudoword and behavioral responses judging correct versus incorrect object-pseudoword pairing via button pressing. We tested two hypotheses: H1: The visual modality will elicit stronger word learning effects than the auditory modality (cf. Bigelow & Poremba, 2014; Cohen et al., 2009; Robinson & Sloutsky, 2004; Thiessen, 2010). H2: The visual modality will benefit from simultaneous presentation and the auditory modality will be better in the sequential presentation (cf. Conway & Christiansen, 2005; Saffran, 2002).

2 Experiment 1: Auditory Sequential

In this experiment, participants were trained with auditory objects followed by spoken pseudowords either in a consistently or inconsistently paired manner. Subsequently, participants were subjected to a testing phase where matched pairs (consistent pairing from the training phase) and violated pairs (mismatched sounds and pseudowords from the consistent pairs) were presented. We predicted that participants would show a difference in ERPs for consistently vs. inconsistently paired items across the training phase and behavioral discrimination and different ERPs for violated vs. matching pairs in the testing phase.

2.1 Methods

2.1.1 Participants

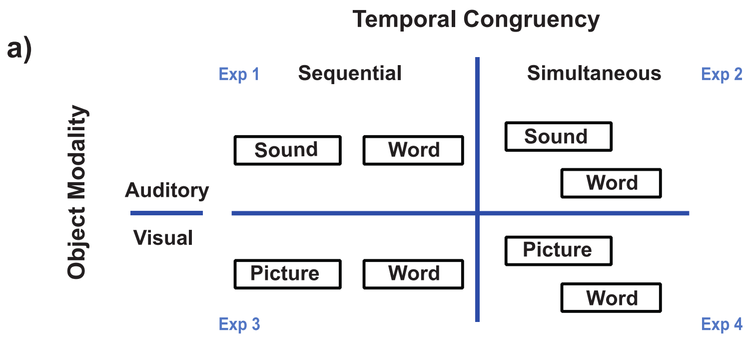
A group of 22 students (16 female), aged 19 to 32 years, were tested (M = 23.86 years; SD = 3.4 years). The participants were all students of the University of Osnabrück, native German speakers, right-handed, reported normal hearing and normal/corrected eye-sight, and no neurological conditions. Written informed consent was given before participation. Participants were compensated for their time either with credit towards their degree programs (mandatory participation in experiments) or 10 euro. All experiments were approved by the ethics committee of the University of Osnabrück and conformed to all aspects of the Declaration of Helsinki.

2.1.2 Stimuli

Auditory object stimuli contained 16 environmental sounds taken from the NESSTI database for environmental sounds (Hocking et al., 2013). The pseudoword stimuli were comprised of 16 disyllabic pseudowords that were derived from existing German nouns. These pseudowords were recorded from a female native German speaker using natural intonation. Each of the auditory object files were cut to 950 ms in length and the pseudoword files were 750 ms long. Each file was saved as a monaural .WAV file, digitized at 44100 Hz, and presented duplicated over two speaker channels at a comfortable volume level that was kept constant across participants. The experiment was presented using Presentation (Neurobehavioral Systems, Inc., Berkeley, CA, USA, version 20.0).

2.1.3 Procedure

Participants were seated in front of a computer monitor with speakers on either side, within reach of two buttons for the behavioral measures. The experiment was divided into a training phase and an immediately following testing phase. The training phase consisted of 128 trials, in which sound-pseudoword combinations were presented. Eight of the 16 auditory objects and pseudowords were allocated to consistent pairings, while the other half was allocated to inconsistent pairings. For the consistent pairing, each sound-word combination was presented eight times (for a total of 64 trials). For the inconsistent pairings, each of the 8 sounds was paired with each of the 8 words once, for a total of 64 trials. In the testing phase, sound-pseudoword pairings from the consistent condition were then presented in a matching or violated manner. The matching pairs consisted of the 8 consistent sound-pseudoword pairs from the training phase, while the 8 violated pairs were recombination of the same items. Because each pair was presented twice, this resulted in 16 pairs per condition and 32 trials overall (see Fig. 1a, for an example of the sound object and pseudoword distribution in the training phase and Fig. 1b, for the distribution in the testing phase). Eight randomization lists were created in order to counterbalance stimuli across consistent and inconsistent conditions, to redistribute sound and pseudoword pairings, and presentation order (i.e., four lists present stimuli in the reverse order). A fixation cross was always present on the screen.



b) Training Phase

	Consistent Pairs		Inconsistent Pairs	
1st - 4th	Object 1	Pseudoword 1	Object 9	Pseudoword 9
	Object 1	Pseudoword 1	Object 10	Pseudoword 9
	Object 1	Pseudoword 1	Object 11	Pseudoword 9
	Object 1	Pseudoword 1	Object 12	Pseudoword 9
5th- 8th	Object 1	Pseudoword 1	Object 13	Pseudoword 9
	Object 1	Pseudoword 1	Object 14	Pseudoword 9
	Object 1	Pseudoword 1	Object 15	Pseudoword 9
	Object 1	Pseudoword 1	Object 16	Pseudoword 9

c) Testing Phase

Matching Pairs		Violated Pairs	
Object 1	Pseudoword 1	Object 1	Pseudoword 3
Object 2	Pseudoword 2	Object 2	Pseudoword 8
Object 3	Pseudoword 3	Object 3	Pseudoword 2
Object 4	Pseudoword 4	Object 4	Pseudoword 1
Object 5	Pseudoword 5	Object 5	Pseudoword 7
Object 6	Pseudoword 6	Object 6	Pseudoword 5
Object 7	Pseudoword 7	Object 7	Pseudoword 4
Object 8	Pseudoword 8	Object 8	Pseudoword 6

Fig. 1. Experimental setup. (a) The 2x2 experimental design illustrates the difference in experimental manipulation (i.e., temporal congruency and object modality) between the four experiments, in which sequentially presented stimuli are separated by 600 ms and simultaneously presented stimuli overlap 500 ms. (b) Example of the presentation of object-pseudoword pairs in consistent and inconsistent conditions of the training phase. Colors used for visual emphasis. (c) Example of the presentation of object-pseudoword pairs in the matching and violated conditions of the testing phase. Colors used for visual emphasis.

The structure of each trial within experiment 1 was identical. Sounds were presented first, followed by a 600 ms inter-stimulus pause, then a pseudoword was presented. Between each trial, there was a 1600 ms inter-trial pause. During the testing phase, behavioral data were also collected, such that 1600 ms after each trial, participants were asked to judge by button press if the presented pairing was correct or incorrect. The assignment of left and right buttons to yes and no answers was counterbalanced across randomization lists.

2.1.4 EEG Processing

All EEG data were collected at the Kindersprachlabor of the University of Osnabrück. The EEG signals were measured using a REFA amplifier (Twente Medical Systems International, Oldenzaal, The Netherlands) with Ag/AgCl electrodes at extended standard 10-20 electrode positions implemented in 64-channel TMSi caps (Twente Medical Systems International, Oldenzaal, The Netherlands). Monopolar EOG electrodes were placed above and below the left eye and on the temples of each eye to measure eye movement-related activity. All impedances were kept below 10 kΩ. EEG data were continuously recorded using the TMSi Polybench Software (Twente Medical Systems International, Oldenzaal, The Netherlands) with a sampling rate of 512 Hz, average reference, and a ground located on the left collarbone.

EEG pre-processing was conducted in MATLAB (The Mathworks Inc., Natick, MA, USA, version 2017a) using EEGLAB (Delorme & Makeig, 2004, version 14_1_1b). In an initial step, a high-pass filter of 1 Hz (-3dB, cutoff frequency of 1.38 Hz) and a low-pass filter of 30 Hz (-3dB, cutoff frequency of 31.16 Hz) were applied to the continuous EEG data. EEG epochs time-locked to the onset of the pseudowords were then created with a length of 1200 ms and a pre-stimulus baseline of 200 ms. After epochization, EEG data were re-referenced to linked mastoids. Artifact rejection was conducted using manual inspection. A semi-automatic independent component analysis (ICA) was applied to each EEG dataset for eye-artifact correction. After artifact correction, the ICA weights of the 1–30 Hz filtered dataset were applied to a dataset with a high-pass filter of 0.3 Hz (-3dB, cutoff frequency of 0.36 Hz) and a low-pass filter of 30 Hz (-3dB, cutoff frequency of 31.16 Hz), and all statistical analyses were conducted using this 0.3–30 Hz dataset.

Subsequently, grand means were created over all participants for each condition in the training phase (1^{st}–4^{th} presentations of consistent pseudowords, 1^{st}–4^{th} presentations of inconsistent pseudowords, 5^{th}–8^{th} presentations of consistent pseudowords, and 5^{th}–8^{th} presentations of inconsistent pseudowords) and in the testing phase (matching pseudowords and violated pseudowords). The final data set for the first half of the training phase consisted of an average of 31.1 trials (97.2%; *SD* = 1.2 trials) for the consistently paired pseudowords and an average of 31.2 trials (97.6%; *SD* = 1.2 trials) for the inconsistently paired. For the second half of the training phase, the final data set consisted of an average of 30.4 trials (95%; *SD* = 1.6 trials) for the consistent pseudowords and an average of 30.6 trials (95.6%; *SD* = 1.5 trials) for the inconsistently paired pseudowords. For the testing phase, the final data set included 15.5 trials (96.9%; *SD* = 0.7 trials) for the matching pseudowords and 15.1 trials (94.6%; *SD* = 1.0 trials) for the violated pseudowords.

2.1.5 Statistical Analysis

For the behavioral results of each experiment, participants' response accuracy was assessed in SPSS (IBM, Armonk, NY, USA, version 26) using a one-sample t-test contrasting correct versus incorrect responses. In order to compare the overall accuracy results of the 2x2 experimental design, a between-subjects ANOVA was applied with the factors modality (auditory and visual) and congruency (simultaneous and sequential) using SPSS. In order to compare the accuracy of all experiments against one another, a Tukey HSD post-hoc test was applied using SPSS (IBM, Armonk, NY, USA). The statistical comparison of all experiments is presented in the last section.

All statistical analyses for the EEG data were conducted in MATLAB (The Mathworks Inc., Natick, MA, USA) using the Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011, version 20170815). A cluster-based permutation analysis was calculated by means of dependent samples t-tests for each sampling point with an alpha threshold fixed at 0.05. In order for a significant sample to be included in the clustering algorithm, the minimum number of neighborhood channels was set to 2. Using the Monte Carlo method (Maris & Oostenveld, 2007), a permutation test was applied with 1000 randomizations and an alpha threshold of 0.05. The reference electrodes and the EOG electrodes were removed from the cluster-based statistical analysis. In order to capture any significant N400-like effects, a time window of 400–800 ms post-stimulus was pre-selected for analysis across all experiments.

For the training phase, cluster-based statistical analyses were applied in order to test for differences between consistent versus inconsistent object-pseudoword pairs in the first half of the training phase (first through fourth presentations of each pseudoword) and in the second half of the training phase (fifth through eighth presentations of each pseudoword). In the testing phase, a separate cluster-based statistical analysis was applied to test for differences between matching pairs and violated pairs.

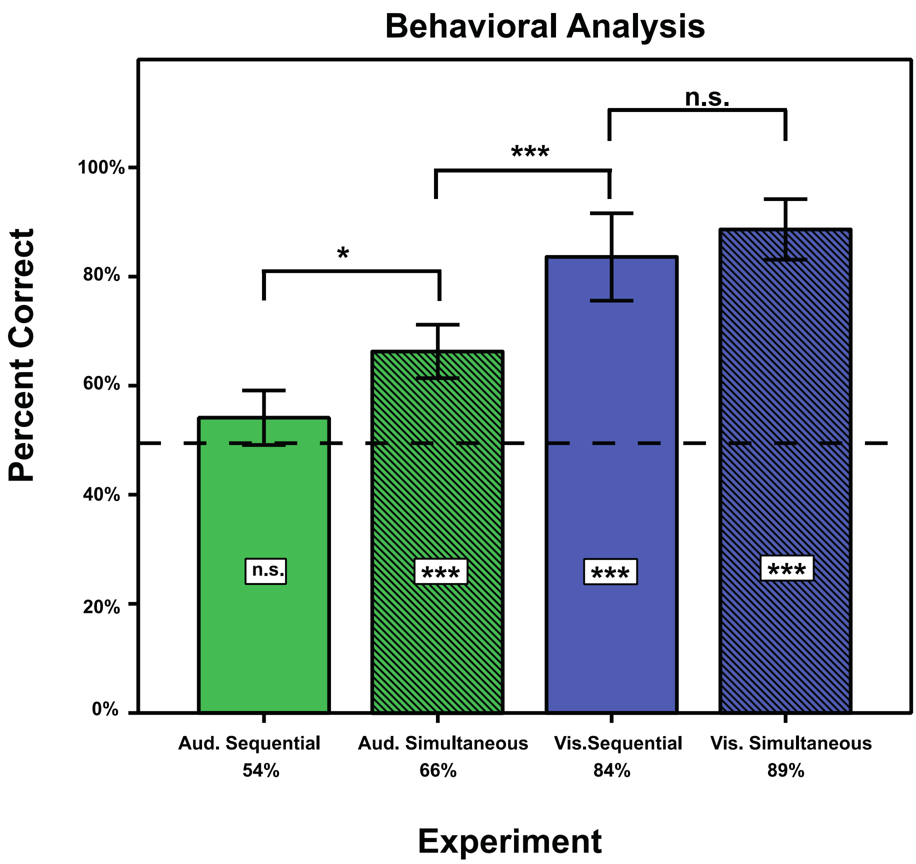


Fig. 2. Behavioral analysis. The average accuracy for participants is depicted per experiment. The dotted line indicates chance level at 50%. The confidence intervals are shown here at 95%. Within each of the bars, it is indicated if the accuracy in each of the experiments was statistically above chance. The paired comparisons between the experiments in order of ascending behavioral performance, as calculated by the Tukey HSD post-hoc analysis, are shown above the bar graph and are presented in the Comparative Results section (n.s. = not significant; * = p < 0.05; *** = p < 0.001).

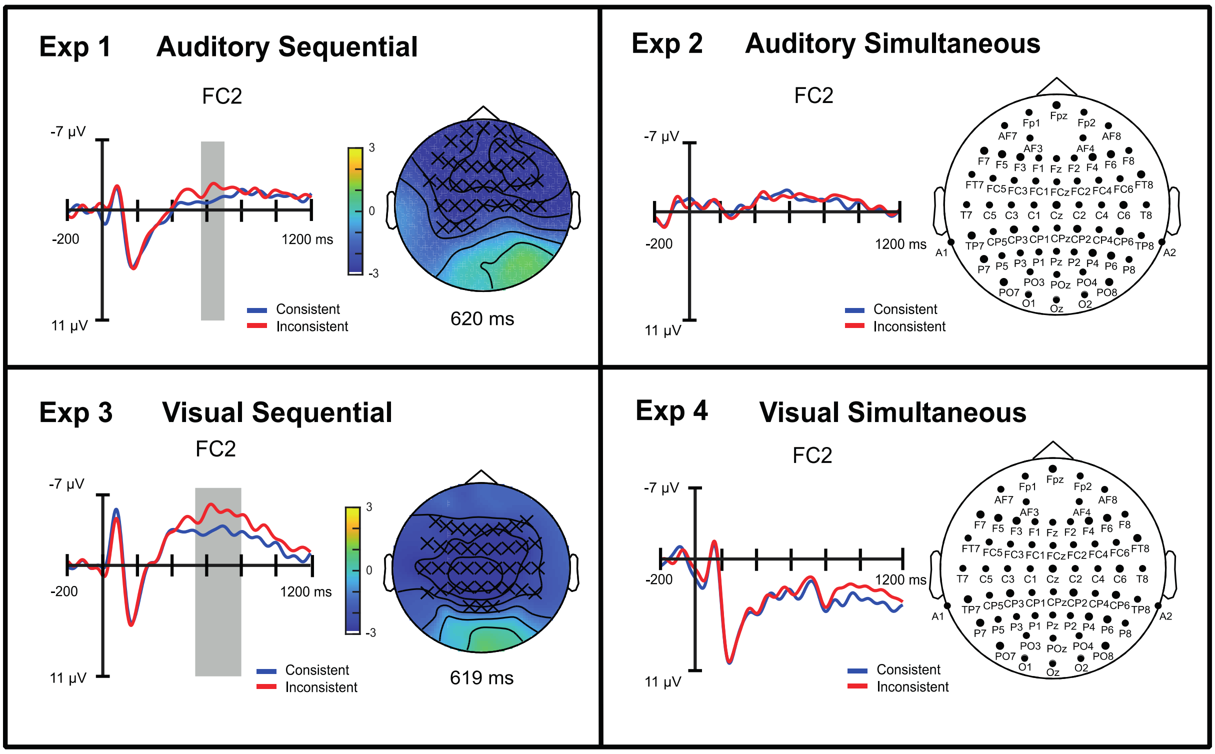


Fig. 3. Training phase results. ERPs time-locked to the onset of the pseudowords in the second half of the training phase are shown for each experiment (10 Hz low-pass filter applied for visualization only). The electrode depicted has been selected as a representative electrode for all experiments. Blue lines represent ERPs to consistent pairs while red lines indicate ERPs to inconsistent pairs. Gray bars indicate time-windows where there is a significant difference between the two conditions. Topological difference maps show peak cluster distribution (the X symbol indicates $p < 0.05$). The head maps indicate electrode positions.

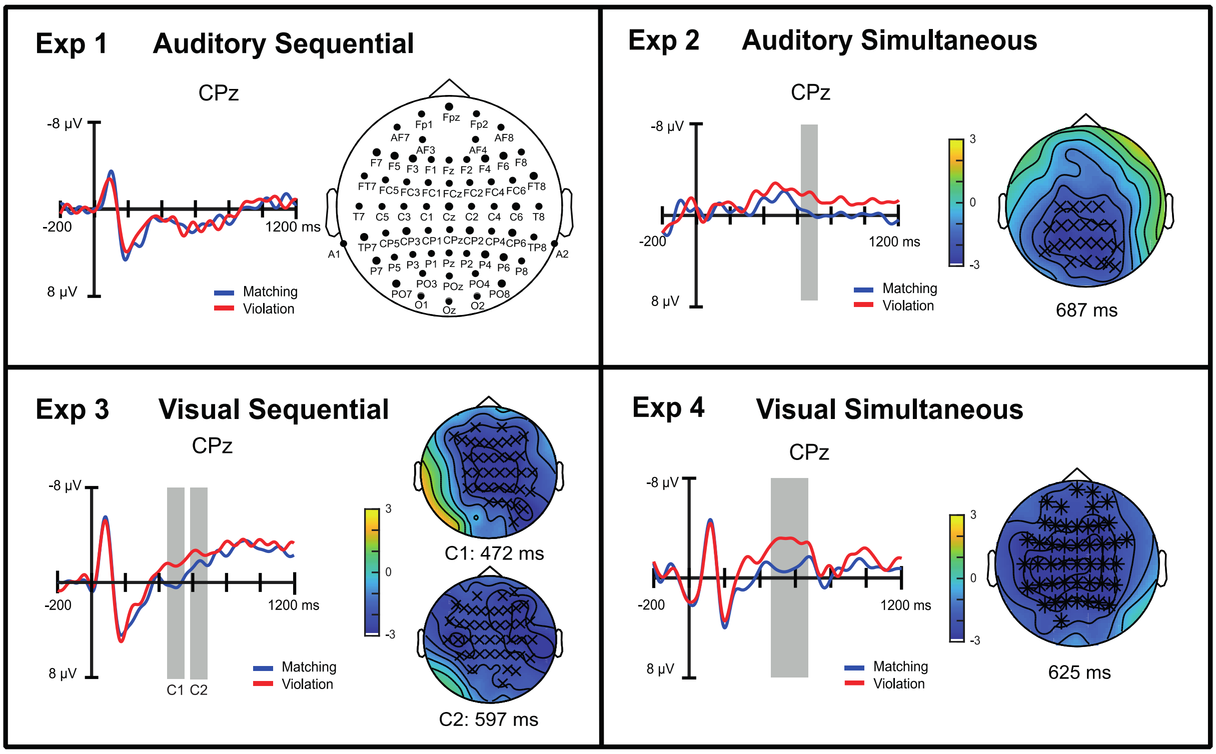


Fig. 4. Testing phase results. The ERPs time-locked to the onset of the pseudowords in the testing phase are shown for each experiment (10 Hz low-pass filter applied for visualization only). The electrode depicted has been selected as a representative electrode for all experiments. Blue lines indicate ERPs to the matched pairs and red lines represent ERPs to the violated pairs. Gray bars indicate time-windows where there is a significant difference between the two conditions (in Experiment 3, C1 refers to cluster 1, while C2 refers to cluster 2). Topological difference maps show peak cluster distribution (the X symbol denotes $p < 0.05$, while the * symbol represents $p < 0.01$). The head map indicates electrode positions.

2.2 Results

2.2.1 Behavioral Results

With only 54% (M = 17.32 correct responses; SD = 3.52) of responses being correct in average, the participants were not above chance ($p > 0.1$) in identifying correct and incorrect sound-pseudoword pairings (see Fig. 2).

2.2.2 ERP Results

In the first half of the training phase, there were no clusters with significant processing differences between inconsistent over consistent object-pseudoword pairing. Yet, there was a significant difference in the second half of the training phase ($p < 0.02$). This difference resulted from more negative ERPs to inconsistent than consistent pairings at fronto-central electrode sites, at 573 – 652 ms after the onset of the pseudoword (see Fig. 3).

In the testing phase, no significant processing differences between the violated and matching pairs were found in the cluster-based analysis (see Fig. 4).

2.2.3. Results summary

Despite showing an emerging N400-like response for inconsistent compared to consistent pairings in the second half of the training phase, no significant ERP or behavioral effects were found in the testing phase. The lack of an N400 lexical expectation effect in the testing phase indicates that these newly formed associations from the training phase were not strong enough to persist even across only a short time period and/or to impact on behavior.

3 Experiment 2: Auditory Simultaneous

Experiment 2 was similar to Experiment 1; however, the auditory stimuli and pseudowords were presented in a simultaneous manner. We predicted that listeners will show less evidence of learning compared to Exp. 1 and thus show neither significant ERP effects nor above chance level behavioral performance.

3.1 Methods

3.1.1 Participants

Twenty-five students participated in this experiment. The final dataset included 24 participants aged 19 to 30 years (11 female; $M = 23.15$ years old; $SD = 3.17$ years). One participant was excluded due to a technical error.

3.1.2 Stimuli and procedure

The auditory object stimuli and the pseudoword stimuli as well as the experimental procedure were the same as in Experiment 1. The current trial construction deviated, however, such that auditory objects were again presented first, but with a 500-ms overlap in presentation with the pseudowords.

3.1.3 EEG processing and statistical analysis

The processing of the EEG data and the statistical analyses were identical to Experiment 1. The final data set for the first half of the training phase consisted of an average of 30.4 trials (95.1%; *SD* = 1.9 trials) for the consistently paired pseudowords and an average of 30.5 trials (95.3%; *SD* = 2.5 trials) for the inconsistently paired ones. For the second half of the training phase, the final data set consisted of an average of 30.3 trials (94.7%; *SD* = 1.7 trials) for the consistent pseudowords and an average of 30.1 trials (94%; *SD* = 2.3 trials) for the inconsistently paired ones. For the testing phase, the final data set included an average of 15 trials (94%; *SD* = 0.9 trials) for the matching pseudowords and an average of 14.5 trials (90.6%; *SD* = 1.5 trials) for the violated pseudowords.

3.2 Results

3.2.1 Behavioral Results

Participants were above chance ($p < 0.001$) in correctly identifying matching and violated sound-pseudoword pairs at 66% ($M = 21.21$ correct responses; $SD = 3.63$) (see Fig. 2).

3.2.2 ERP Results

For the training phase, no significant processing differences were found between consistent and inconsistent sound-word pairs in the first half or in the second half (see Fig. 3).

In the testing phase, the cluster-based permutation test revealed a significant processing difference between violated and matching pairs ($p < 0.05$). The corresponding cluster was observed over centro-parietal regions at 645 – 731 ms after the onset of the pseudoword (see Fig. 4).

3.2.3. Results summary

Although the participants did not show any emerging lexical expectation effects in the training phase, ERP responses in the testing phase did indicate a significant processing difference between 645 –

731 ms for violated over matching pairs. Although this effect occurs rather late for an N400, in the light of participants' above-chance accuracy in behaviorally detecting matching and violated object-pseudoword pairs, this ERP effect is most likely related to a violation of lexical expectation. Thus, adults were able to associate novel words to auditory objects when both were presented simultaneously.

4 Experiment 3: Visual Sequential

In Experiment 3, novel real-world images were presented with spoken pseudowords in a sequential manner, as in Experiment 1. We predicted significant ERP differences for consistently vs. inconsistently paired items across the training phase, and significantly different ERPs for violated vs. matching pairs in the testing phase and behavioral learning effects.

4.1 Methods

4.1.1 Participants

Twenty-six students participated in this experiment. The final dataset included 24 participants (18 female) aged 18 to 27 years (*M* = 22.49 years; *SD* = 2.15 years). One participant was excluded due to an experimenter error and one participant was excluded due to a technical error.

4.1.2 Stimuli and procedure

In this experiment, visual object stimuli were presented instead of auditory objects. Visual objects were taken from the Amsterdam Library of Object Images (Geusebroek et al., 2005) and the Novel Object and Unusual Name database (Horst & Hout, 2016). The spoken pseudowords were the same as in Experiment 1. The procedure also mirrored Experiment 1. Fixation crosses were presented when visual objects were not present.

4.1.3 EEG processing and statistical analysis

The final data set consisted of an average of 31.1 trials (97.1%; *SD* = 1.2 trials) for consistently paired pseudowords in the first half of the training phase and an average of 30.7 trials (96%; *SD* = 1.3 trials) for the inconsistently paired. The second half of the training phase consisted of an average of 30.8

trials (96.1%; *SD* = 1.5 trials) for the consistent pseudowords and an average of 30.6 trials (95.7%; *SD* = 1.5 trials) for the inconsistently paired pseudowords. For the testing phase, the final data set included an average of 15.3 trials (95.8%; *SD* = 1.4 trials) for the matching pseudowords and an average of 15.2 trials (94.8%; *SD* = 1.5 trials) for the violated pseudowords.

4.2 Results

4.2.1 Behavioral Results

With correct responses 84% of the time ($M = 26.75$ correct responses; $SD = 3.01$), participants were above chance in correctly identifying matching and violated picture-pseudoword pairs ($p < 0.001$) (see Fig. 2).

4.2.2 ERP Results

The cluster-based analysis did not reveal any processing differences between consistently paired and inconsistently paired pseudowords in the first half of the training phase. A significant processing difference between inconsistent and consistent pairs was found in the second half of the training phase ($p < 0.02$). This difference occurred between 564 – 800 ms after the onset of the pseudoword, with a central topological distribution at peak activation (see Fig. 3).

In the testing phase, there were two clusters of significant processing differences between violated over matched pairs. Cluster 1 ($p < 0.05$) was found from 451 – 562 ms and cluster 2 ($p < 0.05$) from 582 – 695 ms after the onset of the pseudoword (see Fig. 4). Both clusters showed wide-spread topological distributions during effect peak activation.

4.3.3 Results summary

The ERP analysis of the second half of training phase revealed an emerging lexical expectation effect, suggesting consistent pairings were familiarized over time. This is confirmed by the N400-like effect of violated lexical expectation in the testing phase. The ERP findings are in concord with above chance behavioral performance in which participants were significantly above chance in identifying correct and incorrect object-pseudoword pairings.

5 Experiment 4: Visual Simultaneous

In Experiment 4, participants were presented visual objects and spoken pseudowords in a simultaneous manner. We predicted a processing advantage compared to Experiment 3 and thus larger ERP effects during the training and the testing phase as well as better behavioral performance.

5.1 Methods

5.1.1 Participants

Twenty-two students (13 female), ranging from 19 to 30 years old, participated in this experiment (*M* = 22.65 years; *SD* = 2.99 years).

5.1.2 Stimuli and procedure

The visual objects were the same objects as in Experiment 3. The pseudowords were the same as in all previous experiments. Trial construction mirrored that of Experiment 2. In this experiment, object and word stimuli were presented overlapping. Initially, the object was presented alone for 450 ms. For the remaining 500 ms of the object presentation, the pseudoword was played. After the object has disappeared, a fixation cross was presented until the end of the pseudoword stimulus. Other aspects of the procedure were identical to previous experiments.

4.1.3 EEG processing and statistical analysis

The final data set consisted of an average of 31.2 trials (97.6%; *SD* = 1.0 trials) for consistently paired pseudowords in the first half of the training phase and an average of 30.9 trials (96.6%; *SD* = 1.3 trials) for the inconsistently paired. The second half of the training phase consisted of an average of 31.1 trials (97.2%; *SD* = 1.1 trials) for the consistent pseudowords and an average of 30.7 trials (96%; *SD* = 1.9 trials) for the inconsistently paired pseudowords. For the testing phase, the final data set included an average of 15.5 trials (97.2%; *SD* = 1.1 trials) for the matching pseudowords and an average of 15.1 trials (94.6%; *SD* = 1.3 trials) of the violated pseudowords.

5.2 Results

5.2.1 Behavioral Results

With 88% of the responses being correct (M = 28.36 correct responses; SD = 3.91), the participants were above chance identifying matching and violated picture-pseudoword pairs ($p < 0.001$) (see Fig. 2).

5.2.2 ERP Results

In the training phase, no clusters were found with significant processing differences between inconsistent and consistent pairs in the first or second half (see Fig. 3).

In the testing phase, a single cluster was found with a significant processing difference between violated and matched pairs ($p < 0.01$). This difference was found between 463 – 639 ms after the onset of the pseudoword, with a wide-spread topological distribution in the peak electrode activation of the cluster (see Fig. 4).

5.2.3 Results summary

Despite the ERP analysis of the training phase not yielding any significant effects, the testing phase revealed a significant difference between violated over matching pairs from 463 – 639 ms. Above-chance behavioral responses further indicate that participants were capable of mapping words onto novel visual objects if both were presented in a simultaneous manner.

6 Comparisons between Experiments

Here, statistical analyses including behavioral data from all four experiments are presented. Furthermore, the ERPs from both visual experiments are directly compared in order to test for synchronicity advantages.

6.1 Behavioral Results

For the overall between-subjects ANOVA with the factors modality (visual and auditory) and congruency (simultaneous and sequential), main effects were found for modality $F(1,88) = 78.32$, p

< 0.001, and for congruency $F(1,88) = 8.63$, $p < 0.01$. A Tukey HSD post-hoc analysis was applied to determine if there were differences between the behavioral response accuracy of the individual experiments (see Fig. 2). The Tukey HSD post-hoc test revealed significant differences between the behavioral accuracy of the Auditory Sequential experiment and the Auditory Simultaneous experiment ($p < 0.03$), the Visual Sequential experiment ($p < 0.001$), and the Visual Simultaneous experiment ($p < 0.001$). For the Auditory Simultaneous experiment, differences were found between the Visual Sequential experiment ($p < 0.001$) and the Visual Simultaneous experiment ($p < 0.001$). No significant differences were found between behavioral accuracy of the Visual Sequential experiment and the Visual Simultaneous.

6.2 EEG Results

A difference waveform (violated condition – matched condition) was created from the testing phase of Experiment 3 and Experiment 4. A cluster-based permutation test was calculated by means of independent-samples t-tests with all parameters held the same as in previous analyses. In the 400 - 800 ms time-window, there were no clusters with significant differences between the two testing phases.

6.3 Results summary

The behavioral differences between the individual experiments further support the ERP data in that adult word learning benefits from the visual modality more than from the auditory modality. Within the auditory modality, the comparative behavioral accuracies indicate that adult word learning is induced by the simultaneous presentation of sounds and pseudowords. In the visual modality, both ERP difference wave comparison and behavioral accuracies show that adult word learning does not benefit from one temporal congruency condition over the other.

7 Discussion

Across four experiments, we found evidence for H1, learning object-pseudoword associations in the visual modality is superior to learning in the auditory modality (cf. Bigelow & Poremba, 2014; Cohen et al., 2009; Robinson & Sloutsky, 2004; Thiessen, 2010, for visual advantages in other domains). In contrast, H2, word learning for auditory objects benefits from sequential presentation and word learning for visual objects from simultaneous word presentation (cf. Saffran, 2002), was not confirmed. The detailed training and testing phase results of our experiments will be discussed in turn.

For ERP results of the training phases, only the sequential conditions yielded an emerging N400 indicating processing differences between consistent and inconsistent object-pseudoword pairs. At first glance this seems counterintuitive as both simultaneous conditions, but not both sequential conditions, led to significant results in the testing phases. We argue that this difference in the training phases might be related to the dynamics of how expectations about upcoming stimuli are built up. Recent models of the functional processes reflected in the N400 stress the role of weighted or probabilistic predictions for explaining amplitude differences (Bornkessel-Schlesewsky & Schlesewsky, 2019; Rabovsky, Hansen, & McClelland, 2018). If we assume that learning in the simultaneous condition was facilitated compared to the sequential condition, one could explain the absence of any N400 effects by an efficient trial-to-trial update of the internal predictive model, which may have led to comparable error signals across the consistent and the inconsistent conditions. In the consistent case, participants may have noted the high predictability from the start and successfully predicted the upcoming word. In the inconsistent condition they may have noticed the variation from the start and consequently may have adapted their predictions accordingly. In contrast, if we assume a slower, more difficult learning process for the sequential condition, participants may have built and applied incorrect internal predictive models, specifically for the inconsistent condition, and thus, a difference between consistently and inconsistently paired items emerged. While this explanation is speculative, it remains an interesting observation that the neural system clearly shows a learning effect which is either not stable over time or stays at the implicit level, as it is not linked to subsequent behavioral or testing phase effects.

Consistent with our hypothesized advantage for the visual modality, we found clear N400 effects in the testing phases and superior behavioral performance in both visual conditions. In the auditory modality, learning success was weaker. As such, we conclude that adult word learning benefits from the visual modality compared to auditory. While the current experiments do not allow for conclusions about consolidation and retrieval after longer time periods, the findings are broadly consistent with previously reported visual advantages in adults' learning and perception (Bigelow & Poremba, 2014; Cohen et al., 2009; Robinson & Sloutsky, 2004) and extend those findings to object-word association.

Alternatively, the differences between the auditory and the visual conditions could be explained by unimodal versus multimodal processing effects. It is well documented that multimodal input positively influences processing and memory (cf. Eördegh et al., 2019; Matusz et al., 2015). Thus, we cannot exclude that the observed advantage for the visual domain is rather one for multimodal processing. This would lend support to the idea that language is a multimodal system which integrates visual and auditory information at many levels (cf. Vigliocco, Perniss, & Vinson, 2014) and to the fact that word meanings often relate to categories of objects with multimodal properties or even abstract entities.

Our hypothesis that word learning in the visual modality would benefit from simultaneous stimulus presentation, while the auditory modality would benefit from sequential presentation (cf. Saffran, 2002), was not supported. In the visual modality, no clear benefits of simultaneous presentation over sequential presentation were found. In the auditory modality, the proposed superior sequential condition did not yield any ERP or behavioral effects of successful learning. This is surprising given the consensus in the auditory learning literature that SL, associative learning, and even classical Pavlovian conditioning with auditory stimuli all benefit from sequential presentation (for a review, see Boakes & Costa, 2014). There are several methodological aspects that may have contributed to the unexpected findings in our experiments.

First, the benefit of sequential stimulus presentation from auditory artificial grammar learning experiments (cf. Saffran, 2002) may not extend to auditory associative word learning. Auditory artificial grammars are generally presented in a sequential order, as relations among elements in spoken language are temporally ordered and thus sequential in nature. This is not the case for object labeling, which could benefit from increased associative strength during simultaneous presentation. Second, we speculate that the length of the pause between stimuli in the sequential condition needs to be explored in greater detail. It is known that the greater the temporal separation between two stimuli, the less likely they are to be associated with one another, due to contributions of interference and decay (for a review, see Boakes & Costa, 2014). Thus, the 600 ms-pause between sound and pseudoword in the sequential condition may have been too long for successful association. Future experiments should systematically explore timing as a modulating factor of auditory associative word learning. Third, the particular type of auditory stimuli could have also affected learning in the sequential condition. Where Saffran (2002) presented auditory-verbal stimuli, the current auditory experiments additionally presented auditory-nonverbal stimuli. We speculate that auditory-verbal and auditory-nonverbal short-term memory might be maintained via different mechanisms, for example the phonological loop versus auditory imagery (cf. Baddeley, 2012; Soemer & Saito, 2015), which may have led to difficulties in associating environmental sounds and pseudowords across certain time thresholds. All of these factors cannot be disentangled in the current set of experiments and should be experimentally considered to further understand the mechanisms behind auditory associative word learning.

In sum, the behavioral and ERP results from the current experiments provide evidence that the mechanisms behind associative word learning are affected by both object modality and temporal congruency of stimuli. Adults learn words for visual objects more effectively than for auditory objects in an associative learning paradigm. Simultaneous object-pseudoword presentation seems to strengthen associative learning especially when words for auditory objects were learned. Our findings show how variable word learning can be in the light of different modalities and timing constraints and this may spark further research on how different sensory and also abstract properties of the world are integrated into word meanings.

Part III:
General Discussion

7 General discussion

Across the two papers presented in this cumulative dissertation, the auditory associative word learning abilities of both 10-month-old infants and adults were experimentally investigated using the N400 as a means of both learning as well as violation of expectation effects in a testing phase. The 10- to-12-month-old infants exhibited both a *word-form familiarity* effect and a *pairing consistency effect* in the training phase. In the testing phase, *an N400-like effect* was also observed with a greater negative amplitude for violated sound-word pairs over matching sound-word pairs from 300-400 ms after the onset of the pseudoword. This confirmed all of the hypotheses of the experiment. Not only were infants familiarized with the novel word-forms across the training phase, consistent pairing exposure over time resulted in a reduced amplitude over the inconsistent pairings, indicating that within these consistent pairings, the association between the sounds and pseudowords were becoming stronger across training. In the testing phase, the presence of the N400-like effect indicates that infants were able to recognize violations in these newly learned pairs.

However, the results of the adult experiments were slightly different. First, emerging N400 effects in the training phase were only found in the sequential conditions. These training effects, however, did not translate to violation of lexical-semantic expectation in the testing phase of these conditions. While both the Visual Sequential and Visual Simultaneous conditions produced N400 effects, indicating successful learning, only in the Auditory Simultaneous condition provided evidence that participants were able to recognize violations of consistent novel sound-pseudoword pairs. The behavioral analyses of the testing phases support the ERP findings, as only in those conditions yielding a significant N400 effect also yielded above chance responses in correctly identifying matching and violated object-pseudoword pairs. In light of these results, hypothesis 1, the visual modality will elicit stronger effects compared to the auditory modality, was confirmed in both the ERP and behavioral analyses. However, hypothesis 2, the visual modality will benefit from simultaneous stimulus presentation while the auditory modality will benefit from sequential presentation, was not confirmed. For the visual modality, there was no significant difference between the difference waves (violated – matching) of the ERP waveforms did not significantly differ from one another. The behavioral analysis also did not yield any significant differences in the accuracy ratings between the two temporal conditions. For the auditory modality, no significant effects in the testing phase were found in either the ERP or the behavioral analyses of the sequential condition; yet, the participants in the simultaneous condition were able to successfully identify violations of newly learned sound-pseudoword pairs in both the ERP and the behavioral analyses.

Now that we have seen the word learning abilities of infants and adults in auditory associative word learning experiments, we can compare these abilities across age and modality. In Chapter 4, three hypotheses were presented which help direct the discussion. First, it was predicted that the modality of the object affects word learning differently in infants and adults, specifically pertaining to the visual and auditory modalities (cf. Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003; Thiessen, 2010). Second, it was predicted that the N400 component can be modulated by object modality (cf. Junge et al., 2012; van Petten & Rheinfelder, 1995). Finally, it was predicted that the temporal congruency of stimulus presentation will, in addition to modality, affect word learning (cf. Boakes & Costa, 2014; Conway & Christiansen, 2005; Saffran, 2002; Weike et al., 2007). Each of these hypotheses will be considered and thoroughly discussed in separate sections below. Subsequently, the limitations of the current set of studies will also be presented and discussed. Lastly, we will consider future outlooks for auditory associative word learning.

7.1 Sensory modality and word learning

Within the means of communicating sensory experiences to others and languages having the means for which to do so in the form of lexical items (cf. Chedid et al., 2019; Chen, Zhao, Long, Lu, & Huang, 2019; Lynott, Connell, Brysbaert, Brand, & Carney, 2020; Majid et al., 2018; Miklashevsky, 2018; Speed & Majid, 2017; Vergallito, Petilli, & Marelli, 2020; Winter, Perlman, & Majid, 2018), acquiring the meanings of these words is vital for linguistic development. A majority of research in the field of word learning has largely focused on the visual modality, leaving us with a skewed view of word learning as it pertains to the auditory modality, for example. In the infant study conducted by Cosper and colleagues (2020) was able to show that 10- to-12-month-old infants are capable of mapping novel pseudowords onto auditory objects in an associative learning paradigm in a similar manner as 6- and 14-months-old infants are capable of mapping novel words onto novel visual objects (Friedrich & Friederici, 2008, 2011). Similarly, the set of adult studies conducted by Cosper and colleagues (submitted) found that adults are able to map pseudowords onto visual objects either when the objects and pseudowords are presented simultaneously or sequentially, yet adults are successful at mapping pseudowords onto auditory objects when the stimuli are presented simultaneously, but not sequentially. These results indicate that infants and adults process novel object-word pairs differently between modality; yet, the reasoning for these differences in learning pertaining to the modality of the object are not clear between the studies.

There are several reasons as to why human learners at different ages process objects in various modalities differently. One of the main reasons, as already presented in this dissertation, is how infants and adults process these modalities in unimodal and multimodal situations. Infants are very strong auditory learners, as the auditory modality is not only quite developed at the time of birth, but infants also have auditory input in utero (cf. Banks & Salapatek, 1983; DeCasper & Fifer, 1980; DeCasper & Spence, 1986; Herschkowitz, 1988; Panneton & DeCasper, 1986; Starr et al., 1977; Weitzman & Graziani, 1968). In multimodal situations, infants have been described as having an auditory dominance (Lewkowicz, 1988a, 1988b; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003). Although the study by Cosper and colleagues (2020) was a unimodal auditory study, this auditory dominance may also be able to be extended to unimodal statistical learning. Emberson and colleagues (2019) provided evidence that 8-10-month-old infants are stronger statistical learners in unimodal auditory paradigms than in visual paradigms. These learning differences can also be seen in the N400-like effects found in multimodal audio-visual word learning experiments as compared to unimodal auditory word learning experiments (for a detailed discussion, see Chapter 7.2). However, it is difficult to compare unimodal auditory word learning to multimodal audio-visual word learning as unimodal processing is easier than multimodal processing. In multimodal processing, infants are faced with many challenges including longer habituation phases irrespective of pre-familiarization with the stimuli or the use of novel stimuli (Robinson & Sloutsky, 2010b). Additionally, both familiar unfamiliar auditory stimuli affects the processing of visual stimuli by slowing it down (Robinson et al., 2005; Robinson & Sloutsky, 2007; Sloutsky & Robinson, 2008). Given these studies on multimodal processing, it is a plausible that word learning in multimodal settings may be somewhat slower than in unimodal auditory paradigms. Thus, infants are adept auditory learners and learning words for auditory objects may be more efficient than learning in multimodal situations. Although infants are capable of statistical learning and learning words given their experience with the auditory and the effect auditory processing has on visual processing.

Adults, on the other hand, are very strong statistical learners in multimodal situations, particularly in audio-visual situations. In these multimodal situations, we can see that long-term memory for items learned in a multimodal setting is much stronger than items learned in unimodal auditory or visual situations (Eördegh et al., 2019). Matusz and colleagues (2017) furthermore provide evidence that stimuli in congruent multimodal presentations have an increased memory performance than for stimuli presented in incongruent situations. This could, in part, also have to do with the modality preference of adult processing. In contrast to infants, adults tend to exhibit a visual preference in multimodal situations (Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003). This also can be extended to deficits in auditory processing. Despite immediate recall for auditory statistical learning being stronger than that of visual statistical learning (De Gelder & Vroomen, 1997), short-term memory in the auditory modality is inferior to that of the visual modality (Bigelow & Poremba, 2014). Furthermore, difficulties in the auditory modality, as it pertains to the auditory associative word learning experiments conducted by Cosper and colleagues (submitted), may also stem from the mechanisms vital for maintaining auditory short-term memory. Cosper and colleagues (submitted) presented adults with both auditory-verbal and auditory-nonverbal stimuli. Auditory-verbal stimuli is thought to be internally rehearsed within an articulation-based mechanism known as the phonological loop (Baddeley, 2012), while recent research suggests that auditory-nonverbal stimuli are maintained in short-term memory by a cognitive mechanism related to auditory imagery (Soemer & Saito, 2015). As adults showed difficulties in a sequential auditory associative word learning task (cf. Cosper et al, submitted), it could be that the maintenance of both auditory-verbal and auditory-nonverbal information may interfere with one another if the pause between the sound and the pseudoword is too long (for a detailed discussion on temporal congruency, see Chapter 7.3). Thus, modality of the object seems to affect adult word learning, as multimodal memory is facilitated while a decrease in memory performance is found in the auditory modality.

It is well known that children are able to acquire the meanings of words referring to the auditory modality and that adults also have the ability to use and recognize these words and their meanings in many languages across the globe (cf. Chedid et al., 2019; Chen, Zhao, Long, Lu, & Huang, 2019; Lynott, Connell, Brysbaert, Brand, & Carney, 2020; Majid et al., 2018; Miklashevsky, 2018; Speed & Majid, 2017; Vergallito, Petilli, & Marelli, 2020; Winter, Perlman, & Majid, 2018). The fact that infants and children display an auditory dominance while adults show a visual preference is also known (Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003). This change in modal preference has been suggested to begin around 6 or 7 years of age, or around the time children enter the educational system (Nava & Pavani, 2013). Similarly, Thiessen (2010) described both the developmental constraint to auditory statistical learning as well as the stimulus constraint within adult auditory statistical learning, suggesting that not only does auditory statistical learning vary in infancy and adulthood, but that adults process auditory-verbal and auditory-nonverbal information differently. Cosper and colleagues (2020, submitted) have used these previous findings and applied them to an auditory associative word learning paradigm to directly compare word learning abilities across age and modality, where infant word learning is compared to similar studies and adult word learning is directly compared in a 2x2 experimental design. Not only do infants outperform adults in auditory associative word learning, but adult auditory word learning was not comparable to the more successful multimodal audio-visual word learning. Thus, sensory processing abilities both for the auditory modality and multimodal audio-visual situations affect word learning in humans across the lifespan. Additionally, multimodal word learning may also be affected by modal dominance both in infancy and adulthood, where infants are less successful as compared to auditory unimodal and adults are much more successful word learners in multimodal situations. Therefore, the sensory modality of the object being labelled affects word learning both in infancy and adulthood.

7.2 Sensory modality and the N400

In addition to the sensory modality of the object being labelled affecting word learning, a second hypothesis within the scope of this dissertation predicted that the electrophysiological reflection of word learning, the N400 component, would also be affected and modulated by the modality in which the object is presented. These modulations are predicted to be present in both infants as well as adults and can occur as modulations of amplitude, temporal modulations, and modulations of topological distribution. As previously described, the N400 component is an ERP effect of violated expectation that is linked to lexical-semantic processing and in adults typically peaks around 400 ms after the onset of the target word with a centro-parietal topological distribution (for a detailed review, see Kutas & Federmeier, 2011). In infants, the N400 for word learning is said to typically be delayed (Junge et al., 2012) with reports of time windows stretching from 300 ms after the onset of the target word (Friedrich et al., 2015) to 1200 ms (Torkildsen et al., 2008), indicating a temporal modulation. Thus, it can be seen that age is a modulating factor for the N400 component. Furthermore, we can see that the type of auditory stimuli also has a modulating effect on the N400. The amplitude of the N400 component is influenced by type of stimuli, for example a written word, a spoken word, and a line drawing (for a review, see Kutas & Federmeier, 2000). Friedrich and Friederici (2005) expanded upon this by comparing the ERP waveforms of pseudowords and phonotactically illegal nonwords finding that already at 19 months of age, infants consider pseudowords as potential words but not phonotactically illegal nonwords. Furthermore, van Petten and Rheinfelder (1995) found that the type of auditory stimulus (auditory-verbal vs. auditory-nonverbal) modulated the topological distribution of the N400 effect in that environmental sounds (auditory-nonverbal) elicited N400 effects that were left lateralized. Finally, the amplitude of the N400 effect is also influenced by memory consolidation. Friedrich and colleagues (2015) showed that napping had an effect not only on word learning, but also on the N400 effect in 9-16-month-old infants. Thus, the N400 component can be modulated by many factors. In this subsection, the modulatory effect of object modality will be discussed as it pertains to the N400 effect of violated lexical-semantic expectation in both infants and adults.

In the infant study conducted by Cosper and colleagues (2020) found an N400-like effect of violated expectation in 10- to-12-month-old infants. However, the N400-like effect in the study differed from expectations in two ways. First, the topological distribution of the ERP component varied from the typical centro-parietal distribution pattern found in both infants and adults (for a review, see Kutas & Federmeier, 2011). In the auditory associative word learning paradigm, the N400-like effect found was strongest on the midline, which is to be expected, yet had a waning effect that was left-lateralized. However, this lateralized effect is not completely novel within semantic priming ERP experiments. Adults presented with pairs of word and environmental sound fragments in a paired judgement task were found to have left-lateralized N400 effects for mismatched pairs on the environmental sound (van Petten & Rheinfelder, 1995). Based on these findings, it could be that the topological distribution of the N400-like effect from the infant auditory associative word learning experiment was modulated by the auditory-nonverbal stimuli, also producing a distribution similar to that of adult processing.

Second, the infant N400-like effect found in the auditory associative word learning experiment by Cosper and colleagues (2020) was found much earlier than in the multimodal audio-visual comparisons in 6-month-olds (Friedrich & Friederici, 2011) and 14-month-olds (Friedrich & Friederici, 2008). In 6-month-old infants, an N400-like effect was found between 600 - 900 ms after the onset of the pseudoword in a multimodal associative word learning paradigm (Friedrich & Friederici, 2011). In 14-month-old infants in the same paradigm, the N400-like effect was found at a right parietal cite from 200 - 1000 ms after the onset of the pseudoword and at the parietal midline cite at 600 - 1000 ms (Friedrich & Friederici, 2008). The associative word learning with auditory object stimuli found a wider distribution pattern that was found left lateralized from 300 - 400 ms after the onset of the pseudoword (Cosper et al., 2020). Although the 14-month-olds show an effect from 200 ms, the overlapping cites, in this case the midline, differ between the multimodal studies and the unimodal auditory studies. Junge and colleagues (2012) suggested that the infant N400-like effect may be delayed. With the evidence from multimodal word learning experiments and from the current unimodal auditory word learning experiment combined with the knowledge of sensory processing effects on modality, it is plausible that the infant N400-like effect is modulated by modality and that multimodal word learning experiments delay the onset of the N400 while the auditory modality does not. This, however, is based on a few experiments and would need to be directly compared in order to have a better understanding.

The modal modulation of the N400 effect is not limited to infants. In the adult associative word learning experiments in Cosper and colleagues (submitted), a direct comparison of N400 effects was possible due to the systematic experimental design. In the multimodal design, N400 effects indicating word learning were found both in the sequential and the simultaneous temporal presentation conditions, whereas an N400 effect of lexical-semantic violated expectation indicating successful mapping was only found in the Auditory Simultaneous condition within the auditory modality. The N400 effects of the multimodal conditions were very similar and did not significantly differ from one another. The N400 effect in the Visual Simultaneous condition was found between 463 - 639 ms and in the Visual Sequential condition between 541 - 562 ms and from 582 - 695 ms. In the multimodal conditions, the two temporal presentation conditions differed only slightly; however, the peak cluster distribution was quite broad for both of the temporal conditions. For the Auditory Simultaneous condition, the N400 effect of lexical-semantic violation of expectation was found from 645 -731 ms after the onset of the pseudoword and peak cluster distribution was only found in the centro-parietal region. Not only can we see a delay in the onset of the auditory unimodal associative word learning as compared to multimodal conditions, but we also see a smaller topological distribution in the auditory condition. Thus, the N400 effect can be modulated by modality in an inverse manner as in infants. Here, the auditory modality may delay the onset of the N400 effect and may restrict its topological distribution. Although the current data support these claims, they need to be verified in further testing in order for them to be generalized.

7.3 Temporal congruency of stimulus presentation and word learning

The final hypotheses put forth in this dissertation predicts that, in addition to the modality in which the object is presented, temporal congruency of the presented stimuli affects word learning. This hypotheses is based off of statistical learning paradigms which state that auditory statistical learning benefits from sequential presentation, while visual statistical learning benefits from simultaneous presentation (cf. Conway & Christiansen, 2005; Saffran, 2002). Additionally, classical Pavlovian conditioning also suggests that auditory learning is more successful in trace conditioning (sequential) than in delay conditioning (simultaneous) (cf. Boakes & Costa, 2014; Weike et al., 2007). Cosper and colleagues (2020) presented infants with an auditory associative word learning task with the stimuli presented in a sequential manner. Friedrich and Friederici (2008, 2011) presented 6- and 14-month-old infants with a similar task with visual objects and spoken pseudowords in a simultaneous manner. In both modalities, infants were successful learners. However, neither of these experiments presented multimodal stimuli in a sequential manner or auditory stimuli in a simultaneous manner. Although this is in line with statistical learning predictions (cf. Conway & Christiansen, 2005; Saffran, 2002), and classical conditioning (cf. Boakes & Costa, 2014; Weike et al., 2007), these experiments alone are not sufficient in confirming any benefits within the modalities themselves. Therefore, further research into the benefits of word learning and temporal congruency of stimulus presentation in infant learning is necessary.

However, Cosper and colleagues (submitted) conducted a set of experiments in adult word learning to specifically test the claims of temporal congruency as a modulating factor to word learning. In these experiments, both multimodal and auditory associative word learning were tested with sequential stimulus presentation and simultaneous stimulus presentation in training and testing phases. In the training phases, there were interesting findings across modalities. While both of the sequential conditions elicited emerging N400 effects of semantic priming, neither of the simultaneous conditions presented similar findings. Although these findings may not be related to temporal congruity alone, the underlying cognitive mechanisms which are utilized in these situations seem to be correlated to how the stimuli are presented. Cosper and colleagues (submitted) attributed these differences to the possibility that learners are able to form predictions in sequential presentations, allowing for the testing of hypotheses over time, while the simultaneous conditions may not have allowed for predictions to be formed and the learning process was not based on predictive processing. In the testing phase, the hypotheses that visual statistical learning benefits from simultaneous presentation while auditory statistical learning benefits from sequential presentation were not confirmed. Neither the ERP effects evoked nor the behavioral response accuracies in the visual object conditions were not significantly different from one another, suggesting that there is no clear benefit from either temporal condition. Furthermore, learners in the Auditory Sequential condition were not successful in recognizing violations of consistently paired items, while learners in the Auditory Simultaneous condition were both above chance in behavioral responses and exhibited significant electrophysiological responses to violations of lexical-semantic expectations in familiarized sound-pseudoword pairs. Given that associative word learning is sensitive to the associative strength between objects and their labels, it may be the case that in associative word learning generally benefits from a simultaneous presentation of stimuli. Thus, the temporal congruity of stimulus presentation may have an effect on word learning in an associative word learning paradigm.

7.4 Limitations

This cumulative dissertation has presented, to the author's knowledge, the first studies on auditory associative word learning in both infants and adults. Understanding how we learn words for things we cannot see is vital to understanding how words attributed to various perceptual modalities are learned and how the ability to learn words within a specific modality changes and develops across the span of life. In these two studies, it has become evident that 10- to-12-month-old infants are more than capable of mapping novel words onto real auditory objects in the form of environmental sounds and are able to recognize violations to familiarized sound-word pairs, even after a relatively short training phase (Cosper et al., 2020). Adults, on the other hand, are more limited in their abilities to map novel words onto auditory objects (Cosper et al., submitted). Although adult learners are able to successfully acquire words for environmental sounds when the sound and the word are presented simultaneously, they are not successful learners under the same conditions as infants. Furthermore, while infant word learning abilities are comparable between unimodal auditory and multimodal conditions, adult multimodal word learning far exceeds unimodal auditory word learning. These novel findings have expanded our knowledge on word learning and sensory processing. However, the work presented in the studies by Cosper and colleagues (2020, submitted) only begin to develop the picture as to how words in the auditory modality are learned and how sensory modality affects word learning, how sensory modality modulates the N400, and how the temporal congruity of stimulus presentation affects word learning. This subchapter focuses on the limitations of the findings presented in this dissertation.

First, the experiments conducted on infants are not enough to conclusively determine how auditory associative word learning differs from multimodal audio-visual word learning. Although the experiment from Cosper and colleagues (2020) was based on an associative word learning study utilizing audio-visual input, the designs of the study were not identical (cf. Friedrich & Friederici, 2008, 2011). While the infant study in this dissertation used auditory objects in the form of real environmental sounds, Friedrich and Friederici (2008, 2011) conducted their experiments using digital images of novel, colorful shapes. Moreover, the testing phase of the experiments differed. While the current study intended to test short-term memory immediately after learning, the studies by Friedrich and Friederici (2008, 2011) tested long-term memory with the testing phase occurring 24 hours after the training phase. This, of course, does not allow for a direct comparison between the modalities as it pertains to how effectively infants map novel words onto objects. Additionally, the comparison between the N400-like effects found in each of the studies do not reflect the same processes due to consolidation and memory degradation. In order to compare learning between the modalities in a more direct manner, the study of Cosper and colleagues (2020) should be expanded in order to be able to directly compare learning abilities across the multimodal audio-visual and unimodal auditory modalities. Likewise, the effects of temporal congruity on word learning could not be determined between the studies presented. In order to form a clearer picture, a study similar to Cosper and colleagues (submitted) should be conducted with infants.

Second, all experiments presented in this dissertation used real-word objects, whereas many studies on first language acquisition with use novel visual stimuli. This could be an issue as it is difficult to determine if infants and adults alike had prior knowledge or experience with specific stimuli. These biases or prior knowledge could have led to participants being unable to map a label onto the object as it conflicted with an already existing label. Although all stimuli were surveyed before testing, it is still a possibility that could have skewed the results. The same could be said with the pseudoword stimuli presented. The pseudowords were created based on existing German nouns; however, there were a few reports of participants speaking a dialect of German that either contained an identical lexical item or a similar lexical item. With a participant already having a meaning for a lexical item, the semantic competition could have also hindered the mapping of the pseudoword onto a novel object in any modality. In order to account for both of these possible stimuli biases, the experiments should be replicated with different object and pseudoword stimuli. This would allow for a comparison of results and, in turn, would lead to more generalized implications to be extracted from the findings.

Finally, there are limitations to the generalization of the findings found in the current set of studies as they pertain to word learning, sensory processing, and the perceptual modality norms found within a given language. On generalization to word learning, the studies conducted in Cosper and colleagues (2020, submitted) only present findings of a small population sample. In order to have a generalization on word learning, even within German learners, replications of both studies would need to be conducted, including populations of other age groups and socio-economic groups. The generalization of the current findings as they pertain to sensory processing is intrinsically related to the perceptual modality norms of a given language. Adults have been said to have both a visual preference in multimodal audio-visual processing (cf. Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003) and a better memory for visual objects (Cohen et al., 2009; Yuval-Greenberg & Deouell, 2007, 2009). Furthermore, there is a plethora of evidence that many Indo-European languages, as well as Mandarin Chinese, have more words allocated to the visual modality than to any other modality (cf. Chedid et al., 2019; I.-H. Chen, Zhao, Long, Lu, & Huang, 2019; Lynott, Connell, Brysbaert, Brand, & Carney, 2020; Majid et al., 2018; Miklashevsky, 2018; Speed & Majid, 2017; Vergallito, Petilli, & Marelli, 2020; Winter, Perlman, & Majid, 2018). These findings seem to complement one another, as the sensory processing abilities of individuals with a matured language system seem to process visual information more efficiently and happen to have more words allocated to that particular modality. However, Majid and colleagues (2018) have shown that having a higher codability for words pertaining to the visual modality is not a universal across all languages in the world. As many of the languages presented in the study by Majid and colleagues (2018) are minority languages, it is possible that research is biased to studying cognitive abilities in those populations which speak majority languages, many displaying the visual modality bias. Moreover, the research in sensory modal dominance is also conducted in these majority language populations. Thus, further research is necessary in order to determine if the visual dominance in adulthood is a universal dominance across human cognition or if it is a product of first-word education systems and/or a byproduct of perceptual modality norms within certain language populations. These limitations to the current set of studies do not undermine the efficacy of their implications, but rather are meant to keep the generalizations extrapolated within the scope of the study designs and participant groups tested.

7.5 Outlook

The studies presented in this dissertation have provided initial evidence about infant and adult auditory associative word learning and the effects of sensory modality on word learning, on the N400 component, and the effect of the temporal congruity of stimulus presentation on word learning (2020, submitted). In order to have a more complete understanding about how we learn words for auditory objects, more research is needed. Although these studies have assessed the abilities of 10- to-12-month-old infants, a greater understanding is needed at other milestones in infancy and childhood. Furthermore, the current studies have focused on healthy populations with participants able to perceive all sensory modalities. However, there are individual differences in other population groups that would extend the current findings to include populations with other dominant senses, for example congenitally blind participants. Blind individuals have a different sensory experience with the world and may exhibit greater learning capabilities in auditory associative word learning. Additionally, other special populations with enhanced auditory expertise may also elicit stronger learning effects, for example professional musicians. Sensory processing can vary between individuals and expanding the current paradigms to include these individuals would allow for more generalized implications on human modality specific word learning. Our understanding of language learning has not reached its limit and the implications of this dissertation show that there is still much to be discovered. Sensory processing and word learning are intertwined and considering both together will bring a greater understanding of human communication and learning abilities.

8 References

Abel, A. D., Schneider, J., & Maguire, M. J. (2018). N400 Response Indexes Word Learning from Linguistic Context in Children. *Language Learning and Development*, *14*(1), 61–71. https://doi.org/10.1080/15475441.2017.1362347

Aguilar, J. M., & Plante, E. (2014). Learning of Grammar-Like Visual Sequences by Adults With and Without Language-Learning Disabilities. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1394–1404. https://doi.org/10.1044/2014_JSLHR-L-13-0124

Angwin, A. J., Phua, B., & Copland, D. A. (2014). Using semantics to enhance new word learning: An ERP investigation. *Neuropsychologia*, *59*, 169–178. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2014.05.002

Aslin, R. N. (2017). Statistical learning: a powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science*. Wiley-Blackwell. https://doi.org/10.1002/wcs.1373

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, *63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: Lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia*, *79*, 33–41. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2015.10.020

Balass, M., Nelson, J. R., & Perfetti, C. A. (2010). Word learning: An ERP investigation of word experience effects on recognition and word processing. *Contemporary Educational Psychology*, *35*(2), 126–140. https://doi.org/10.1016/J.CEDPSYCH.2010.04.001

Banks, M. S., & Salapatek, P. (1983). Infant visual perception. In P. H. Mussen, M. M. Haith, & J. J. Campos (Eds.), *Handbook of child psychology: Vol. 2. Infancy and developmental psychobiology* (pp. 435–571). New York: Wiley.

Bentin, S., Kutas, M., & Hillyard, S. A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology*, *30*(2), 161–169. https://doi.org/10.1111/j.1469-8986.1993.tb01729.x

Bentin, S., Kutas, M., & Hillyard, S. A. (1995). Semantic processing and memory for attended and unattended words in dichotic listening: Behavioral and electrophysiological evidence. *Journal* *of Experimental Psychology: Human Perception and Performance*, *21*(1), 54–67. https://doi.org/10.1037/0096-1523.21.1.54

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*(4), 343–355. https://doi.org/10.1016/0013-4694(85)90008-2

Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(49), 12916–12921. https://doi.org/10.1073/pnas.1712966114

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Bermúdez-Margaretto, B., Beltrán, D., Cuetos, F., & Domínguez, A. (2018). Brain Signatures of New (Pseudo-) Words: Visual Repetition in Associative and Non-associative Contexts. *Frontiers in Human Neuroscience*, *12*, 354. https://doi.org/10.3389/fnhum.2018.00354

Bigelow, J., & Poremba, A. (2014). Achilles' ear? Inferior human short-term and recognition memory in the auditory modality. *PLoS ONE*, *9*(2), e89914. https://doi.org/10.1371/journal.pone.0089914

Bloom, P. (2000). *How Children Learn the Meanings of Words*. *How Children Learn the Meanings of Words*. Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/3577.001.0001

Boakes, R. A., & Costa, D. S. J. (2014). Temporal contiguity in associative learning: Interference and decay from an historical perspective. *Journal of Experimental Psychology*. *Animal Learning and Cognition*. https://doi.org/10.1037/xan0000040

Boddy, J. (1986). Event-related potentials in chronometric analysis of primed word recognition with different stimulus onset asynchronies. *Psychophysiology*, *23*(2), 232–245. https://doi.org/10.1111/j.1469-8986.1986.tb00624.x

Bomba, P. C., & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology*, *35*(2), 294–328. https://doi.org/10.1016/0022-0965(83)90085-1

Borgström, K., Torkildsen, J. von K., & Lindgren, M. (2015a). Event-related potentials during word mapping to object shape predict toddlers' vocabulary size. *Frontiers in Psychology*, *6*(FEB), 1–15. https://doi.org/10.3389/fpsyg.2015.00143

Borgström, K., Torkildsen, J. von K., & Lindgren, M. (2015b). Substantial gains in word learning ability between 20 and 24 months: A longitudinal ERP study. *Brain and Language*, *149*, 33–45. https://doi.org/10.1016/j.bandl.2015.07.002

Borgström, K., Torkildsen, J. von K., & Lindgren, M. (2016). Visual event-related potentials to novel objects predict rapid word learning ability in 20-month-olds. *Developmental Neuropsychology*, *41*(5–8), 308–323. https://doi.org/10.1080/87565641.2016.1243111

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*, *10*(FEB), 298. https://doi.org/10.3389/fpsyg.2019.00298

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development*, *8*(3), 278–302. https://doi.org/10.1080/15475441.2011.614893

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*(4), 298–304. https://doi.org/10.1111/j.0956-7976.2005.01531.x

Bowerman, M., & Choi, S. (2001). Shaping meaning for language: Universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language Acquisition and Conceptual Development* (pp. 475–511). Cambridge: Cambridge University Press.

Braddick, O., & Atkinson, J. (2011). Development of human visual function. *Vision Research*. Pergamon. https://doi.org/10.1016/j.visres.2011.02.018

Braun, M., Jacobs, A. M., Hahne, A., Ricker, B., Hofmann, M., & Hutzler, F. (2006). Model-generated lexical activity predicts graded ERP amplitudes in lexical decision. *Brain Research*, *1073–1074*, 431–439. https://doi.org/10.1016/J.BRAINRES.2005.12.078

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A.-F., Sommer, J., Wolbers, T., & Knecht, S. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage*, *25*(3), 958–968. https://doi.org/10.1016/J.NEUROIMAGE.2004.12.019

Bruner, J. S. (1983). *Child's Talk: Learning to use a Language*. New York, NY: W. W. Norton & Company.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127–132. https://doi.org/10.1016/j.cognition.2011.06.010

Chedid, G., Brambati, S. M., Bedetti, C., Rey, A. E., Wilson, M. A., & Vallet, G. T. (2019). Visual and auditory perceptual strength norms for 3,596 French nouns and their relationship with other psycholinguistic variables. *Behavior Research Methods*, 1–12. https://doi.org/10.3758/s13428-019-01254-w

Chen, I.-H., Zhao, Q., Long, Y., Lu, Q., & Huang, C.-R. (2019). Mandarin Chinese modality exclusivity norms. *PLOS ONE*, *14*(2), e0211336. https://doi.org/10.1371/journal.pone.0211336

Chen, S., Wang, L., & Yang, Y. (2014). Acquiring concepts and features of novel words by two types of learning: Direct mapping and inference. *Neuropsychologia*, *56*(1), 204–218. https://doi.org/10.1016/j.neuropsychologia.2014.01.012

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160055. https://doi.org/10.1098/rstb.2016.0055

Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(14), 6008–6010. https://doi.org/10.1073/pnas.0811884106

Colombo, J. (2001). The Development of Visual Attention in Infancy. *Annual Review of Psychology*, *52*(1), 337–367. https://doi.org/10.1146/annurev.psych.52.1.337

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, *112*(2), 347–382. https://doi.org/10.1037/0033-295X.112.2.347

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Cosper, S. H., Männel, C., & Mueller, J. L. (2020). In the absence of visual input: Electrophysiological evidence of infants' mapping of labels onto auditory objects. *Developmental Cognitive Neuroscience*, *45*, 100821. https://doi.org/10.1016/j.dcn.2020.100821

Cosper, S. H., Männel, C., & Mueller, J. L. (submitted). Mechanisms of word learning: Benefits from the visual modality and synchronicity of labeled objects.

Cummings, A., Čeponiene, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research*, *1115*(1), 92–107. https://doi.org/10.1016/j.brainres.2006.07.050

Cunillera, T., Toro, J. M., Sebastian-Galles, N., & Rodriguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research*, *1123*(1), 168–178. https://doi.org/10.1016/J.BRAINRES.2006.09.046

De Gelder, B., & Vroomen, J. (1997). Modality Effects in Immediate Recall of Verbal and Non-verbal Information. *European Journal of Cognitive Psychology*, *9*(1), 97–110. https://doi.org/10.1080/713752541

de Haan, M., Johnson, M. H., Maurer, D., & Perrett, D. I. (2001). Recognition of individual faces and average face prototypes by 1- and 3-month-old infants. *Cognitive Development*, *16*(2), 659–678. https://doi.org/10.1016/S0885-2014(01)00051-X

de Haan, M., & Nelson, C. A. (1997). Recognition of the Mother's Face by Six-Month-Old Infants: A Neurobehavioral Study. *Child Development*, *68*(2), 187–210. https://doi.org/10.1111/j.1467-8624.1997.tb01935.x

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176. https://doi.org/10.1126/SCIENCE.7375928

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, *9*(2), 133–150. https://doi.org/10.1016/0163-6383(86)90025-1

DeLoache, J. S., Strauss, M. S., & Maynard, J. (1979). Picture perception in infancy. *Infant Behavior and Development*, *2*(1), 77–89. https://doi.org/10.1016/S0163-6383(79)80010-7

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/J.JNEUMETH.2003.10.009

Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behavior and Development*, *5*(2–4), 261–276. https://doi.org/10.1016/S0163-6383(82)80036-2

Demany, L., McKenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature*, *266*(5604), 718–719. https://doi.org/10.1038/266718a0

Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research*. Elsevier.

https://doi.org/10.1016/j.brainres.2008.03.071

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*(4), 547–582. https://doi.org/10.1111/j.1551-6709.2009.01023.x

Emberson, L. L., Misyak, J. B., Schwade, J. A., Christiansen, M. H., & Goldstein, M. H. (2019). Comparing statistical learning across perceptual modalities in infancy: An investigation of underlying learning mechanism(s). *Developmental Science*, *22*(6). https://doi.org/10.1111/desc.12847

Eördegh, G., Őze, A., Bodosi, B., Puszta, A., Pertich, Á., Rosu, A., … Nagy, A. (2019). Multisensory guided associative learning in healthy humans. *PLOS ONE*, *14*(3), e0213094. https://doi.org/10.1371/journal.pone.0213094

Eswaran, H., Lowery, C. L., Wilson, J. D., Murphy, P., & Preissl, H. (2004). Functional development of the visual system in human fetus using magnetoencephalography. *Experimental Neurology*, *190*(SUPPL. 1), 52–58. https://doi.org/10.1016/j.expneurol.2004.04.007

Field, T. M., Cohen, D., Garcia, R., & Greenberg, R. (1984). Mother-stranger face discrimination by the newborn. *Infant Behavior and Development*, *7*(1), 19–25. https://doi.org/10.1016/S0163-6383(84)80019-3

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467. https://doi.org/10.1037/0278-7393.28.3.458

François, C., Cunillera, T., Garcia, E., Laine, M., & Rodriguez-Fornells, A. (2017). Neurophysiological evidence for the interplay of speech segmentation and word-referent mapping during novel word learning. *Neuropsychologia*, *98*, 56–67. https://doi.org/10.1016/j.neuropsychologia.2016.10.006

Friedrich, M., & Friederici, A. D. (2005). Phonotactic Knowledge and Lexical-Semantic Processing in One-year-olds: Brain Responses to Words and Nonsense Words in Picture Contexts. *Journal of Cognitive Neuroscience*, *17*(11), 1785–1802. https://doi.org/10.1162/089892905774589172

Friedrich, M., & Friederici, A. D. (2008). Neurophysiological correlates of online word learning in 14-month-old infants. *Neuroreport, 19*(18), 1757–1761. https://doi.org/Doi 10.1097/Wnr.0b013e328318f014

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

Friedrich, M., & Friederici, A. D. (2017). The origins of word learning: Brain responses of 3-month-olds indicate their rapid association of objects and words. *Developmental Science*, *20*(2). https://doi.org/10.1111/desc.12357

Friedrich, M., Wilhelm, I., Born, J., & Friederici, A. D. (2015). Generalization of word meanings during infant sleep. *Nature Communications*, *6*(1), 6004. https://doi.org/10.1038/ncomms7004

Friedrich, M., Wilhelm, I., Mölle, M., Born, J., & Friederici, A. D. (2017a). The sleeping infant brain anticipates development. *Current Biology*, *27*(15), 2374-2380.e3. https://doi.org/10.1016/j.cub.2017.06.070

Friedrich, M., Wilhelm, I., Mölle, M., Born, J., & Friederici, A. D. (2017b). The sleeping infant brain anticipates development. *Current Biology*, 2374–2380. https://doi.org/10.1016/j.cub.2017.06.070

Gerken, L. (1996). Phonological and distributional information in syntax acquisition. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 411–425). Mahwah, NJ: Lawrence Erlbaum.

Geusebroek, J.-M., Burghouts, G. J., & Smeulders, A. W. M. (2005). The Amsterdam Library of Object Images. *International Journal of Computer Vision*, *61*(1), 103–112. https://doi.org/10.1023/B:VISI.0000042993.50813.60

Gleitman, L. (1990). The Structural Sources of Verb Meanings. *Language Acquisition*, *1*(1), 3–55. https://doi.org/10.1207/s15327817la0101_2

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Baby Wordsmith. *Current Directions in Psychological Science*, *15*(1), 30–33. https://doi.org/10.1111/j.0963-7214.2006.00401.x

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95–112. https://doi.org/10.1007/BF02289823

Heikkilä, J., Alho, K., Hyvönen, H., & Tiippana, K. (2015). Audiovisual Semantic Congruency

During Encoding Enhances Memory Performance. *Http://Dx.Doi.Org/10.1027/1618-3169/A000279*. https://doi.org/10.1027/1618-3169/a000279

Herschkowitz, N. (1988). Brain development in the fetus, neonate and infant. *Neonatology*, *54*(1), 1–19. https://doi.org/10.1159/000242818

Hocking, J., Dzafic, I., Kazovsky, M., & Copland, D. A. (2013). NESSTI: Norms for Environmental Sound Stimuli. *PLoS ONE*, *8*(9), e73382. https://doi.org/10.1371/journal.pone.0073382

Höhle, B. (2009). Bootstrapping mechanisms in first language acquisition. *Linguistics*, *47*(2), 359–382. https://doi.org/10.1515/LING.2009.013

Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional Elements in Infants' Speech Processing: The Role of Determiners in the Syntactic Categorization of Lexical Elements. *Infancy*, *5*(3), 341–353. https://doi.org/10.1207/s15327078in0503_5

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language*, *35*(1), 66–85. https://doi.org/10.1016/0093-934X(88)90101-0

Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, *14*(6), 938–950. https://doi.org/10.1162/089892902760191153

Hollich, G. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). Breaking the language barrier: An emergentist coalition model for the origins of language learning. *Monographs of the Society for Research in Child Development*, *65*(3). https://doi.org/10.1111/1540-5834.00090

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3

Horst, J. S., & Samuelson, L. K. (2008). Fast Mapping but Poor Retention by 24-Month-Old Infants. *Infancy*, *13*(2), 128–157. https://doi.org/10.1080/15250000701795598

Johnson, E. K. (2016). Constructing a Proto-Lexicon: An Integrative View of Infant Language Development. *Annual Review of Linguistics*, *2*(1), 391–412. https://doi.org/10.1146/annurev-linguistics-011415-040616

Johnson, E. K., Seidl, A., & Tyler, M. D. (2014). The edge factor in early word segmentation: Utterance-level prosody enables word Form Extraction by 6-month-olds. *PLoS ONE*, *9*(1).

https://doi.org/10.1371/journal.pone.0083546

Jones, S. S., & Smith, L. B. (2002). How children know the relevant properties for generalizing object names. *Developmental Science, 5*(2), 219–232. https://doi.org/10.1111/1467-7687.00224

Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development*, *62*(3), 499–516. https://doi.org/10.1111/j.1467-8624.1991.tb01547.x

Joseph, S., Kumar, S., Husain, M., & Griffiths, T. D. (2015). Auditory working memory for objects vs. features. *Frontiers in Neuroscience, 9*, 13. https://doi.org/10.3389/fnins.2015.00013

Junge, C., Cutler, A., & Hagoort, P. (2012). Electrophysiological evidence of early word learning. *Neuropsychologia*, *50*(14), 3702–3712. https://doi.org/10.1016/j.neuropsychologia.2012.10.012

Junge, C., Cutler, A., & Hagoort, P. (2014). Successful word recognition by 10-month-olds given continuous speech both at initial exposure and test. *Infancy*, *19*(2), 179–193. https://doi.org/10.1111/infa.12040

Jusczyk, P. W. (2002). How infants adapt speech-processing capacities to native-language structure. *Current Directions in Psychological Science*, *11*(1), 15–18. https://doi.org/10.1111/1467-8721.00159

Kaczer, L., Bavassi, L., Petroni, A., Fernández, R. S., Laurino, J., Degiorgi, S., … Pedreira, M. E. (2018). Contrasting dynamics of memory consolidation for novel word forms and meanings revealed by behavioral and neurophysiological markers. *Neuropsychologia*, *117*, 472–482. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2018.07.001

Key, A. P. F., Molfese, D. L., & Ratajczak, E. D. (2006). ERP Indicators of Learning in Adults. *Developmental Neuropsychology*, *29*(2), 379–395. https://doi.org/10.1207/s15326942dn2902_5

Kiefer, M. (2002). The N400 is modulated by unconsciously perceived masked words: further evidence for an automatic spreading activation account of N400 priming effects. *Cognitive Brain Research, 13*(1), 27–39. https://doi.org/10.1016/S0926-6410(01)00085-4

Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: four markers for a link between auditory and conceptual brain systems. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(47), 12224–12230.

https://doi.org/10.1523/JNEUROSCI.3579-08.2008

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. https://doi.org/10.1016/S0010-0277(02)00004-5

Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., … Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*(3), 220–224. https://doi.org/10.1111/1467-9280.02435

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, *7*(3), 302–307. https://doi.org/10.1038/nn1197

Kooijman, V., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, *24*(1), 109–116. https://doi.org/https://doi.org/10.1016/j.cogbrainres.2004.12.009

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470. https://doi.org/10.1016/S1364-6613(00)01560-6

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science (New York, N.Y.)*, *207*(4427), 203–205. https://doi.org/10.1126/SCIENCE.7350657

Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes*. Hillsdale, NJ: Erlbaum.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 Effects of Prediction from Association in Single-word Contexts. *Journal of Cognitive Neuroscience*, *25*(3), 484–502. https://doi.org/10.1162/jocn_a_00328

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Lecanuet, J. P., & Schaal, B. (1996). Fetal sensory competencies. *European Journal of Obstetrics* *and Gynecology and Reproductive Biology*. Elsevier Ireland Ltd. https://doi.org/10.1016/0301-2115(96)02509-2

Lewkowicz, D. J. (1985). Bisensory Response to Temporal Frequency in 4-Month-Old Infants. *Developmental Psychology*, *21*(2), 306–317. https://doi.org/10.1037/0012-1649.21.2.306

Lewkowicz, D. J. (1988a). Sensory dominance in infants: I. Six-month-old infants' response to auditory-visual compounds. *Developmental Psychology*, *24*(2), 155–171.

Lewkowicz, D. J. (1988b). Sensory dominance in infants: II. Ten-month-old infants' response to auditory-visual compounds. *Developmental Psychology*, *24*(2), 172–182. https://doi.org/10.1037/0012-1649.24.2.172

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., … Levinson, S. C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(45), 11369–11376. https://doi.org/10.1073/pnas.1720419115

Männel, C., & Friederici, A. D. (2013). Accentuate or repeat? Brain signatures of developmental periods in infant word recognition. *Cortex*, *49*(10), 2788–2798. https://doi.org/10.1016/J.CORTEX.2013.09.003

Männel, C., Schaadt, G., Illner, F. K., van der Meer, E., & Friederici, A. D. (2017). Phonological abilities in literacy-impaired children: Brain potentials reveal deficient phoneme discrimination, but intact prosodic processing. *Developmental Cognitive Neuroscience*, *23*, 14–25. https://doi.org/10.1016/J.DCN.2016.11.007

Männel, C., Schipke, C. S., & Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds. *Developmental Cognitive Neuroscience*, *5*, 86–94. https://doi.org/10.1016/J.DCN.2013.01.003

Maouene, J., Hidaka, S., & Smith, L. B. (2008). Body parts and early-learned verbs. *Cognitive Science: A Multidisciplinary Journal*, *32*(7), 1200–1216. https://doi.org/10.1080/03640210802019997

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data.

Journal of Neuroscience Methods, *164*(1), 177–190. https://doi.org/10.1016/J.JNEUMETH.2007.03.024

Matusz, P. J., Thelen, A., Amrein, S., Geiser, E., Anken, J., & Murray, M. M. (2015). The role of auditory cortices in the retrieval of single-trial auditory-visual object memories. *European Journal of Neuroscience*, *41*(5), 699–708. https://doi.org/10.1111/ejn.12804

Matusz, P. J., Wallace, M. T., & Murray, M. M. (2017). A multisensory perspective on object memory. *Neuropsychologia*, *105*, 243–252. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2017.04.008

Mayor, J., & Plunkett, K. (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychological Review*, *117*(1), 1–31. https://doi.org/10.1037/a0018130

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872

Mestres-Misse, A., Rodriguez-Fornells, A., & Munte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex*, *17*(8), 1858–1866. https://doi.org/10.1093/cercor/bhl094

Miklashevsky, A. (2018). Perceptual experience norms for 506 Russian nouns: Modality rating, spatial localization, manipulability, imageability and other variables. *Journal of Psycholinguistic Research*, *47*(3), 641–661. https://doi.org/10.1007/s10936-017-9548-1

Miller, T. M., Schmidt, T. T., Blankenburg, F., & Pulvermüller, F. (2018). Verbal labels facilitate tactile perception. *Cognition*, *171*, 172–179. https://doi.org/10.1016/J.COGNITION.2017.10.010

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience*, *5*(3), 317–334. https://doi.org/10.1162/jocn.1993.5.3.317

Mills, D. L., Plunkett, K., Prat, C., & Schafer, G. (2005). Watching the infant brain learn words: effects of vocabulary size and experience. *Cognitive Development*, *20*(1), 19–31. https://doi.org/10.1016/J.COGDEV.2004.07.001

Mills, D. L., Prat, C., Zangl, R., Stager, C. L., Neville, H. J., & Werker, J. F. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *Journal of Cognitive Neuroscience*, *16*(8), 1452–1464. https://doi.org/10.1162/0898929042304697

Minagawa-Kawai, Y., van der Lely, H., Ramus, F., Sato, Y., Mazuka, R., & Dupoux, E. (2011). Optical Brain Imaging Reveals General Auditory and Language-Specific Processing in Early Infant Development. *Cerebral Cortex*, *21*(2), 254–261. https://doi.org/10.1093/cercor/bhq082

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, *26*(4), 393–424. https://doi.org/10.1207/s15516709cog2604_1

Naigles, L. (1990). Children Use Syntax To Learn Verb Meanings. *Journal of Child Language*, *17*(2), 357–374. https://doi.org/10.1017/S0305000900013817

Namy, L. L. (2012). Getting Specific: Early General Mechanisms Give Rise to Domain-Specific Expertise in Word Learning. *Language Learning and Development*, *8*(1), 47–60. https://doi.org/10.1080/15475441.2011.617235

Nava, E., & Pavani, F. (2013). Changes in Sensory Dominance During Childhood: Converging Evidence From the Colavita Effect and the Sound-Induced Flash Illusion. *Child Development*, *84*(2), 604–616. https://doi.org/10.1111/j.1467-8624.2012.01856.x

Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: two modes of word acquisition? *Developmental Science*, *6*(2), 136–142. https://doi.org/10.1111/1467-7687.00263

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756–766. https://doi.org/10.1037/0096-1523.24.3.756

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by english-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, *43*(1), 1–19. https://doi.org/10.1006/jmla.2000.2698

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., … Von Grebmer Zu Wolfsthurn, S. (2020). Dissociable effects of prediction and integration during language comprehension: evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1791), 20180522. https://doi.org/10.1098/rstb.2018.0522

Nobre, A. C., & McCarthy, G. (1994). Language-Related ERPs: Scalp Distributions and Modulation by Word Type and Semantic Priming. *Journal of Cognitive Neuroscience*, *6*(3), 233–255. https://doi.org/10.1162/jocn.1994.6.3.233

Nobre, A. C., & McCarthy, G. (1995). Language-related field potentials in the anterior-medial temporal lobe: II. Effects of word type and semantic priming. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *15*(2), 1090–1098. https://doi.org/10.1523/JNEUROSCI.15-02-01090.1995

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Obrig, H., Mock, J., Stephan, F., Richter, M., Vignotto, M., & Rossi, S. (2017). Impact of associative word learning on phonotactic processing in 6-month-old infants: A combined EEG and fNIRS study. *Developmental Cognitive Neuroscience*, *25*, 185–197. https://doi.org/10.1016/J.DCN.2016.09.001

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869

Ortu, D., Allan, K., & Donaldson, D. I. (2013). Is the N400 effect a neurophysiological index of associative relationships? *Neuropsychologia*, *51*(9), 1742–1748. https://doi.org/10.1016/j.neuropsychologia.2013.05.003

Oviatt, S., & Cohen, P. (2000). Perceptual user interfaces: multimodal interfaces that process what comes naturally. *Communications of the ACM*, *43*(3), 45–53. https://doi.org/10.1145/330534.330538

Panneton, R. K., & DeCasper, A. J. (1986). Newborns' postnatal preference far a prenatally experienced melody. In *Paper presented at the International Conference on Infant Studies*. Beverly Hills, CA.

Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science*, *23*(7), 728–733. https://doi.org/10.1177/0956797612438734

Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumit, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(20), 11702–11705. https://doi.org/10.1073/pnas.1934290100

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 399–411). Hillsdale, NJ: Lawrence Erlbaum.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information Processing and Cognition: The Loyota Symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Preissl, H., Lowery, C. L., & Eswaran, H. (2005). Fetal Magnetoencephalography: Viewing the Developing Brain In Utero. *International Review of Neurobiology*. Academic Press Inc. https://doi.org/10.1016/S0074-7742(05)68001-4

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The Birth of Words: Ten-Month-Olds Learn Words Through Perceptual Salience. *Child Development*, *77*(2), 266–280. https://doi.org/10.1111/j.1467-8624.2006.00869.x

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Quinn, P. C., Slater, A. M., Brown, E., & Hayes, R. A. (2001). Developmental change in form categorization in early infancy. *British Journal of Developmental Psychology*, *19*(2), 207–218. https://doi.org/10.1348/026151001166038

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, *2*(9), 693–705. https://doi.org/10.1038/s41562-018-0406-4

Raviv, L., & Arnon, I. (2018). The developmental trajectory of children’s auditory and visual statistical learning abilities: modality-based differences in the effect of age. *Developmental Science*, *21*(4), e12593. https://doi.org/10.1111/desc.12593

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, *29*(6), 819–865. https://doi.org/10.1207/s15516709cog0000_31

Reynolds, G. D. (2015). Infant visual attention and object recognition. *Behavioural Brain Research*, *285*, 34–43. https://doi.org/10.1016/j.bbr.2015.01.015

Reynolds, G. D., Courage, M. L., & Richards, J. E. (2013). The Development of Attention. In D. Reisberg (Ed.), *The Oxford Handbook of Cognitive Psychology* (pp. 1000–1013). New York, NY: Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780195376746.013.0063

Rhodes, S. M., & Donaldson, D. I. (2007). Association and not semantic relationships elicit the N400 effect: Electrophysiological evidence from an explicit language comprehension task. *Psychophysiology*, *0*(0), 070914092401002-??? https://doi.org/10.1111/j.1469-8986.2007.00598.x

Richards, D. D., & Goldfarb, J. (1986). The episodic memory model of conceptual development: An integrative viewpoint. *Cognitive Development*, *1*(3), 183–219. https://doi.org/10.1016/S0885-2014(86)80001-6

Robinson, C. W., Howard, E. M., & Sloutsky, V. M. (2005). Mechanisms underlying the effects of labels on cognitive development. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *27*(27).

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development*, *75*(5), 1387–1401. https://doi.org/10.1111/j.1467-8624.2004.00747.x

Robinson, C. W., & Sloutsky, V. M. (2007). Visual processing speed: effects of auditory input on visual processing. *Developmental Science*, *10*(6), 734–740. https://doi.org/10.1111/j.1467-7687.2007.00627.x

Robinson, C. W., & Sloutsky, V. M. (2010a). Development of cross-modal processing. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 135–141. https://doi.org/10.1002/wcs.12

Robinson, C. W., & Sloutsky, V. M. (2010b). Effects of multimodal presentation and stimulus familiarity on auditory and visual processing. *Journal of Experimental Child Psychology*, *107*(3), 351–358. https://doi.org/10.1016/j.jecp.2010.04.006

Robinson, C. W., & Sloutsky, V. M. (2019). Two mechanisms underlying auditory dominance: Overshadowing and response competition. *Journal of Experimental Child Psychology*, *178*, 317–340. https://doi.org/10.1016/j.jecp.2018.10.001

Rossi, S., Hartmüller, T., Vignotto, M., & Obrig, H. (2013). Electrophysiological evidence for modulation of lexical processing after repetitive exposure to foreign phonotactic rules. *Brain and Language*, *127*(3), 404–414. https://doi.org/10.1016/J.BANDL.2013.02.009

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*(1), 172–196. https://doi.org/10.1006/JMLA.2001.2839

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Samuelson, L. K., & McMurray, B. (2017). What does it take to learn a word? *Wiley Interdisciplinary Reviews. Cognitive Science*, *8*(1–2). https://doi.org/10.1002/wcs.1421

Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development*, *69*(1), 94–104. https://doi.org/10.1111/j.1467-8624.1998.tb06136.x

Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, *73*(1), 1–33. https://doi.org/10.1016/S0010-0277(99)00034-7

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*, *5*(7), 700–703. https://doi.org/10.1038/nn873

Sarter, N. B. (2006). Multimodal information presentation: Design guidance and research challenges. *International Journal of Industrial Ergonomics*, *36*(5), 439–445. https://doi.org/10.1016/j.ergon.2006.01.007

Schiavo, J. K., & Froemke, R. C. (2019). Capacities and neural mechanisms for auditory statistical learning across species. *Hearing Research*. Elsevier B.V. https://doi.org/10.1016/j.heares.2019.02.002

Schmidt, T. T., Miller, T. M., Blankenburg, F., & Pulvermüller, F. (2019). Neuronal correlates of label facilitated tactile perception. *Scientific Reports*, *9*(1), 1606. https://doi.org/10.1038/s41598-018-37877-w

Shtyrov, Y. (2011). Fast mapping of novel word forms traced neurophysiologically. *Frontiers in Psychology*, *2*(NOV), 1–9. https://doi.org/10.3389/fpsyg.2011.00340

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6038–6043. https://doi.org/10.1073/pnas.1017617108

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1-2 SPEC. ISS.), 39–91. https://doi.org/10.1016/s0010-0277(96)00728-7

Slobin, D. I. (2001). Form-function relations: how do children find out what they are? In M. Bowerman & S. C. Levinson (Eds.), *Language Acquisition and Conceptual Development* (pp. 406–449). Cambridge: Cambridge University Press.

Sloutsky, V. M., & Fisher, A. V. (2011). *The Development of Categorization*. *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 54). https://doi.org/10.1016/B978-0-12-385527-5.00005-X

Sloutsky, V. M., & Napolitano, A. C. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development*, *74*(3), 822–833. https://doi.org/10.1111/1467-8624.00570

Sloutsky, V. M., & Robinson, C. W. (2008). The role of words and sounds in infants' visual processing: From overshadowing to attentional tuning. *Cognitive Science*, *32*(2), 342–365. https://doi.org/10.1080/03640210701863495

Sloutsky, V. M., Yim, H., Yao, X., & Dennis, S. (2017). An associative account of the development of word learning. *Cognitive Psychology*, *97*, 1–30. https://doi.org/10.1016/J.COGPSYCH.2017.06.001

Smith, L. B., Jayaraman, S., Clerkin, E. M., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, *22*(4), 325–336. https://doi.org/10.1016/J.TICS.2018.02.004

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: a dumb attentional mechanism? *Cognition*, *60*(2), 143–171. https://doi.org/10.1016/0010-0277(96)00709-3

Smith, L. B., Maouene, J., & Hidaka, S. (2007). The body and children's word learning. In J. M. Plumert & J. P. Spencer (Eds.), *The emerging spatial mind* (pp. 168–192).

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Snyder, J. S., & Gregg, M. K. (2011). Memory for sound, with an ear toward hearing in complex auditory scenes. *Attention, Perception, & Psychophysics*, *73*(7), 1993–2007. https://doi.org/10.3758/s13414-011-0189-4

Soemer, A., & Saito, S. (2015). Maintenance of auditory-nonverbal information in working memory. *Psychonomic Bulletin and Review*, *22*(6), 1777–1783. https://doi.org/10.3758/s13423-015-0854-z

Speed, L. J., & Majid, A. (2017). Dutch modality exclusivity norms: Simulating perceptual modality in space. *Behavior Research Methods*, *49*(6), 2204–2218. https://doi.org/10.3758/s13428-017-0852-3

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology*, *15*(6), 626–636. https://doi.org/10.1037/0012-1649.15.6.626

Spelke, E. S. (1981). The infant's acquisition of knowledge of bimodally specified events. *Journal of Experimental Child Psychology*, *31*(2), 279–299. https://doi.org/10.1016/0022-0965(81)90018-7

Spelke, E. S., Smith Born, W., & Chu, F. (1983). Perception of moving, sounding objects by four-month-old infants. *Perception*, *12*(6), 719–732. https://doi.org/10.1068/p120719

Starr, A., Amlie, R. N., Martin, W. H., & Sanders, S. (1977). Development of auditory function in newborn infants revealed by auditory brainstem potentials. *Pediatrics*, *60*(6).

Tabullo, A., Yorio, A., Zanutto, S., & Wainselboim, A. (2015). ERP correlates of priming in language and stimulus equivalence: Evidence of similar N400 effects in absence of semantic content. *International Journal of Psychophysiology*, *96*(2), 74–83. https://doi.org/10.1016/j.ijpsycho.2015.03.004

Taxitari, L., Twomey, K. E., Westermann, G., & Mani, N. (2019). The Limits of Infants' Early Word Learning. *Language Learning and Development*, 1–21. https://doi.org/10.1080/15475441.2019.1670184

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*(1), 21. https://doi.org/10.1186/1471-2202-10-21

Teixidó, M., François, C., Bosch, L., & Männel, C. (2019). The role of prosody in early speech segmentation and word-referent mapping: Electrophysiological evidence. In P. P. Fabra & N. Esteve-Gibert (Eds.), *The Development of Prosody in First Language Acquisition.* (pp. 79–100). Amsterdam: John Benjamins.

Thierry, G., Vihman, M., & Roberts, M. (2003). Familiar words capture the attention of 11-month-olds in less than 250 ms. *NeuroReport*, *14*(18). https://doi.org/10.1097/00001756-200312190-00004

Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cognitive Science*, *34*(6), 1093–1106. https://doi.org/10.1111/j.1551-6709.2010.01118.x

Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160056. https://doi.org/10.1098/rstb.2016.0056

Tincoff, R., & 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

Tincoff, R., & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, *17*(4), 432–444. https://doi.org/10.1111/j.1532-7078.2011.00084.x

Tomasello, M. (2000). The social-pragmatic theory of word learning. *PragmaticsPragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, *10*(4), 401–413. https://doi.org/10.1075/prag.10.4.01tom

Torkildsen, J. von K., Friis Hansen, H., Svangstu, J. M., Smith, L., Simonsen, H. G., Moen, I., & Lindgren, M. (2009). Brain dynamics of word familiarization in 20-month-olds: Effects of productive vocabulary size. *Brain and Language*, *108*(2), 73–88. https://doi.org/10.1016/J.BANDL.2008.09.005

Torkildsen, J. von K., Svangstu, J. M., Hansen, H. F., Smith, L., Simonsen, H. G., Moen, I., & Lindgren, M. (2008). Productive vocabulary size predicts event-related potential correlates of fast mapping in 20-month-olds. *Journal of Cognitive Neuroscience*, *20*(7), 1266–1282. https://doi.org/10.1162/jocn.2008.20087

Torkildsen, J. von K., Syversen, G., Simonsen, H. G., Moen, I., & Lindgren, M. (2007a). Electrophysiological correlates of auditory semantic priming in 24-month-olds. *Journal of Neurolinguistics*, *20*(4), 332–351. https://doi.org/10.1016/J.JNEUROLING.2007.02.003

Torkildsen, J. von K., Syversen, G., Simonsen, H. G., Moen, I., & Lindgren, M. (2007b). Electrophysiological correlates of auditory semantic priming in 24-month-olds. *Journal of Neurolinguistics*, *20*(4), 332–351. https://doi.org/10.1016/J.JNEUROLING.2007.02.003

van Petten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, *33*(4), 485–508. https://doi.org/10.1016/0028-3932(94)00133-A

Vergallito, A., Petilli, M. A., & Marelli, M. (2020). Perceptual modality norms for 1,121 Italian words: A comparison with concreteness and imageability scores and an analysis of their impact in word processing tasks. *Behavior Research Methods*, 1–18. https://doi.org/10.3758/s13428-019-01337-8

Vigliocco, G., Kousta, S.-T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: The role of emotion. *Cerebral Cortex*, *24*(7), 1767–1777. https://doi.org/10.1093/cercor/bht025

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130292. https://doi.org/10.1098/rstb.2013.0292

Walker-Andrews, A. S. (1986). Intermodal Perception of Expressive Behaviors. Relation of Eye and Voice? *Developmental Psychology*, *22*(3), 373–377. https://doi.org/10.1037/0012-1649.22.3.373

Waxman, S. R., & Booth, A. E. (2001). Seeing Pink Elephants: Fourteen-Month-Olds' Interpretations of Novel Nouns and Adjectives. *Cognitive Psychology*, *43*(3), 217–242. https://doi.org/10.1006/cogp.2001.0764

Weike, A. I., Schupp, H. T., & Hamm, A. O. (2007). Fear acquisition requires awareness in trace but not delay conditioning. *Psychophysiology*, *44*(1), 170–180. https://doi.org/10.1111/j.1469-8986.2006.00469.x

Weitzman, E. D., & Graziani, L. J. (1968). Maturation and topography of the auditory evoked response of the prematurely born infant. *Developmental Psychobiology*, *1*(2), 79–89. https://doi.org/10.1002/dev.420010203

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology*, *34*(6), 1289–1309. https://doi.org/10.1037/0012-1649.34.6.1289

Wildt, E., Rohlfing, K. J., & Scharlau, I. (2019). The Role of Saliency in Learning First Words. *Frontiers in Psychology*, *10*(MAY), 1150. https://doi.org/10.3389/fpsyg.2019.01150

Winter, B., Perlman, M., & Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition*, *179*, 213–220. https://doi.org/10.1016/j.cognition.2018.05.008

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272. https://doi.org/10.1037/0033-295X.114.2.245

Yu, C., & Smith, L. B. (2007). Rapid Word Learning Under Uncertainty via Cross-Situational Statistics. *Psychological Science*, *18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262. https://doi.org/10.1016/J.COGNITION.2012.06.016

Yu, C., Zhong, Y., & Fricker, D. (2012). Selective Attention in Cross-Situational Statistical Learning: Evidence From Eye Tracking. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00148

Yuval-Greenberg, S., & Deouell, L. Y. (2007). What you see is not (always) what you hear: Induced gamma band responses reflect cross-modal interactions in familiar object recognition. *Journal of Neuroscience*, *27*(5), 1090–1096. https://doi.org/10.1523/JNEUROSCI.4828-06.2007

Yuval-Greenberg, S., & Deouell, L. Y. (2009). The dog's meow: Asymmetrical interaction in cross-modal object recognition. *Experimental Brain Research*, *193*(4), 603–614. https://doi.org/10.1007/s00221-008-1664-6