Title
Systematic Review of the Effects of Sleep on Memory and Word Learning in Infancy

Journal
Language Learning, 73(2)

ISSN
0023-8333

Authors
Belia, Margherita
Keren-Portnoy, Tamar
Vihman, Marilyn

Publication Date
2023-06-01

DOI
10.1111/lang.12544

Title: Systematic review of the effects of sleep on memory and word learning in infancy.

Authors' names (affiliation): 1) Margherita Belia (University of York), 2) Tamar Keren-Portnoy (University of York), 3) Marilyn Vihman (University of York).

Acknowledgements: We would like to thank Prof. Guillaume Thierry for his insightful comments on a preliminary version of the paper.

Declaration of interests: None

Complete address of corresponding author:
Margherita Belia
Department of Language and Linguistic Science
University of York
Heslington
York
YO10 5DD
United Kingdom

E-mail address of the corresponding author: mb1915@york.ac.uk

Systematic review of the effects of sleep on memory and word learning in infancy.

Abstract

This systematic review aims to map current research on the associations between sleep and the memory processes involved in word learning in infancy. Only 16 studies were found to address this topic directly, identifying associations between infant sleep and the consolidation of phonological structure as well as the consolidation and generalisation of new word form-meaning associations. Some studies investigated changes in brain responses after word learning and in sleep parameters during post-learning sleep. Others investigated the longitudinal effects of sleeping patterns on later vocabulary development. All but one of these studies identified positive associations between sleep and word learning in early childhood, thus extending similar findings in adults and school-aged children. However, several gaps in current research on early lexical development and sleep are identified; we argue that future investigations should address these gaps to better understand vocabulary development and to bridge the gap between memory and language acquisition research.

Keywords: sleep, word learning, infants, memory, generalization, language development

Funding sources: This work is supported by the UK Economic and Social Research Council (ESRC). The funding source had no specific role in study design; in the collection, analysis and interpretation of data; in the writing of the report; in the decision to submit the article for publication.

1. Introduction

In the first years of life, children acquire new words and language skills at an impressively faster rate than adults who learn a foreign language. In the same period, children’s sleeping times are much longer and their sleep is considerably deeper than that of older children, adolescents and adults (Galland, Taylor, Elder & Herbison, 2012). It is now well-established in the adult literature that sleep plays an active, causative and beneficial role in the formation, strengthening and enhancement of memory for newly learned words (e.g. Dumay & Gaskell, 2007; Tamminen, Payne, Stickgold, Wamsley & Gaskell, 2010). Evidence for such effects on word learning in children is also rapidly accumulating in the case of both nocturnal and daytime sleep (see Axelsson, Williams & Horst, 2016, for a review). However, the effects of sleep in early language development, i.e. in children under the age of 3, remains largely unexplored.

Memory formation is generally thought to develop as a three-step process, comprising encoding, consolidation and retrieval (Wojcik, 2013). Encoding occurs when a first memory representation for the experience of an event or a stimulus is created immediately after the experience itself. However, this representation is usually weak, tends to decay over time, and can be overwritten by new experiences. Memory consolidation is thought to be the process by which the newly formed representations are strengthened and become progressively more resistant to interference and decay (Seehagen, Zmyj & Herbert, 2019). Memory consolidation for individually encoded memories is seen as resulting from synaptic consolidation (Born, Rasch & Gais, 2006), a covert memory process occurring without additional experience of the stimulus. It consists in the strengthening and remodelling of synaptic connections between the neurons, or long-term potentiation (LTP), that, when fired, constitute the item memory representation, supported by increased efficacy in the collective firing of synapses. Besides being consolidated, new neural representations are also integrated into existing neural networks, which undergo reorganisation to include the newly integrated information. This process, known as system consolidation, also occurs off-line, i.e., without additional experiences of the stimulus (Born, Rasch & Gais, 2006). The third step of memory formation as described in Wojcik (2013) is retrieval, which consists in the “reactivation of a memory trace” (p.2). Retrieval is the process whereby a memory is recalled and it is also crucial in the long-term retention of a memory trace, as every time it is recalled it is also reactivated and that contributes to its strengthening.

The complementary learning systems (CLS) account, a dual architecture memory system model (McClelland, McNaughton, O’ Reilly, 1995; Davis & Gaskell, 2009), provides an explanation as to how these processes may co-exist. This model was developed to explain the simultaneously plastic and stable nature of the memory system, which must be amenable to change following the acquisition of new knowledge, but at the same time stable enough to prevent this new knowledge from overwriting or interfering with what is already known (a phenomenon known as ‘catastrophic interference’: McClelland et al., 1995; McClelland, 2013). The CLS model describes learning as resulting from the action of two memory systems, operating at different rates: the neocortex and the medial temporal lobes (MTL), especially the hippocampus. The neocortex includes the sensorimotor areas of the brain and hosts our semantic long-term (LT) memory system. Indeed, it is where structured knowledge is elaborated and processed, in the form of representations mapped onto sensorimotor areas. New information is integrated into the existing network, which may undergo some reorganisation. Abstractions based on prior experiences result from the overlapping nature of representations in the knowledge network. These integration processes take place over a relatively prolonged period of time, which ensures that the new information being acquired is successfully integrated into the network without deleting existing knowledge. On the other hand, the hippocampus is essential at the beginning of the learning process as a way of encoding the to-be-learned information or experience as it occurs. In fact, the role of the hippocampal region is to rapidly encode and link cortical patterns associated with a multifaceted experience into a single memory trace in LT memory, which then becomes a memory for a unique episode (i.e. the other major type of long-term memory, episodic memory). In addition, to use the newly encoded memories in a flexible way, the hippocampus cooperates with other brain regions, especially the frontal lobes (McClelland et al., 1995). However, these memory traces endure in the hippocampus for a limited period of time only before they decay; they must be reinstated in the neocortex to be consolidated and remembered.

The Active System Consolidation (ASC) model (Born & Wilhelm, 2012) is proposed as a way to explain the consolidation process and the role of sleep in it. It is based on the tenet that hippocampal memory reactivation during sleep is the fundamental mechanism underlying memory consolidation. According to this model, memory traces for experienced events are rapidly assembled in the hippocampus and also partly in the neocortex. While the hippocampal representations are more complete but less stable, encoded material in the neocortex provides a more stable trace but is less complete. During subsequent sleep and, especially, Slow Wave Sleep (SWS), the newly acquired memories are jointly and repeatedly reactivated in both hippocampal and neocortical networks. During this process, the hippocampal traces function as blueprints for the strengthening of the corresponding neocortical representations, which thereby become gradually redistributed in long-term neocortical networks (and accordingly less dependent on the hippocampus).

The specific electrophysiological state of the brain during this process is the key determinant for consolidation. Indeed, slow-waves sleep (SWS) seems to create the optimal conditions for the replay of episodic memory traces in the hippocampus and their subsequent redistribution in neocortical networks (Diekelmann & Born, 2010). SWS takes its name from the slow oscillations specifically generated in the neocortex during this sleep stage. The fundamental function of neocortical slow oscillations is to temporally bind together the firing activity and silent states of neurons in the hippocampus and the neocortex, by synchronising the oscillatory activity of the brain regions involved in memory consolidation, i.e. the thalamo-cortical networks and the hippocampus. Crucially, these two brain regions each produce distinct oscillatory activity, i.e. thalamo-cortical spindles and hippocampal sharp-wave ripples, the latter corresponding to instances of memory reactivation in the hippocampus (James, Gaskell, Weighall & Henderson, 2017).

More specifically, by synchronizing thalamo-cortical spindles and hippocampal sharp-wave ripples, the slow oscillations generated in the neocortex produce spindle-ripple events which become associated with memory reactivations in the hippocampus. Thus, spindle-ripple events act as a mediating mechanism that enables the activation of specific hippocampal and neocortical neuronal networks, corresponding to memory traces, at approximately the same time. In other words, the synchronizing effect of slow wave oscillation, via the formation of spindle-ripple events, ensures that the memory reactivation in the hippocampus and neocortex are synchronized. As a result, neocortical memory representations are progressively strengthened as the corresponding hippocampal memories are reactivated, and hippocampal memories are gradually redistributed to the more stable long-term networks supporting structured knowledge in the neocortex. The importance of SWS and spindles in memory is supported by findings in which, in adults, the intensity of sleep spindle activity during SWS has repeatedly been found to be positively correlated with memory performance after sleep (Gais, Mölle, Helms & Born, 2002; Tamminen et al., 2010).

Although research on the effects of sleep on memory and language development in infants and young children is in its early days, evidence confirming sleep-associated benefits in several memory consolidation processes is mounting. For example, sleep has been found to promote infant memory retention (i.e. recognition or recall of the specific items presented prior to sleep), as measured in recall and imitation of actions observed prior to sleep (Seehagen, Konrad, Herbert & Schneider, 2015). Infant memory generalisation (i.e. the extraction of core characteristics from a set of several items) has also been found to be improved by sleep, e.g. as observed in infants' ability to imitate the specific actions performed in a learning phase with items that are different from but similar to them (Konrad, Seehagen, Schneider & Herbert, 2016). Generalisation moves knowledge beyond retention, by providing the fundamental resource to transcend the particular events and items experienced and create categories. Crucially, the same processes underlying memory for actions have also been observed in other aspects of cognition and, importantly, language learning.

Indeed, the distinction between memory retention and generalisation is also meaningful for studying how language acquisition and in particular word learning occur when form and meaning are encoded together in the same episodic trace. In the case of word learning, retention corresponds to veridical memory for the actual word form-meaning associations previously learned, e.g., for the label of a referent and the specific referent experienced. Veridical memory can also be formed for unimodal stimuli, such as word forms not yet linked to any meaning. On the other hand, generalisation is considered to result from the extraction of invariant features among several items (Rasch & Born, 2013).

Like veridical memory in language acquisition, generalisation has been studied by looking at both word forms alone and word-meaning pairs. As regards word forms, Houston and Jusczyk (2003), for example, investigated whether infants were able to recognise newly taught word forms after a one-day delay, when uttered by the same speaker as in the familiarisation phase or by a different speaker. Recognising the form in a novel voice can be considered an example of generalisation of the specific sound patterns constituting the phonological form of a word. In fact, this process entails extracting invariant auditory features from a set of individually heard instances of a particular word. Although they did not focus on the role of nocturnal sleep between familiarisation and testing, Houston and Jusczyk found that 7.5-month-olds were able to recognise familiarised word forms uttered by a novel speaker at the delayed test only when primed by hearing words in the voice used during familiarisation too. Houston & Jusczyk (2000) observed that 10.5-month olds, but not 7.5-month olds, could recognise newly familiarised words when produced at test, with no delay, by speakers of the opposite sex. These findings suggest that infants’ word form generalisation abilities develop with age, and that they entail generalising over memory traces which contain highly detailed and even non-linguistic (indexical) information (Foulkes & Docherty, 2006). As regards generalisation of word meanings, exposure to multiple, different but similar referents associated with the same label appear to reinforce the core of a memory trace in semantic memory via repeated reactivation of the recently encoded hippocampal memory, while weakening the activation of inconsistent details, allowing the brain to extract invariant patterns and develop a schematic and general representation (Rasch & Born, 2013). This process is called multi-item generalisation (Stickgold & Walker, 2013), i.e. the constitution of schemas emerging after the reactivation and subsequent reorganization of new and current memory representations during sleep (Lewis & Durrant, 2011). In this way, previously unseen exemplars can be recognised as belonging to the category named by that label.

The aim of this systematic review is to outline the memory mechanisms underlying word learning in early infancy and the effects of sleep on those memory processes by synthesizing and analysing the findings of the relevant literature. In particular, the review will seek to answer the following questions: What evidence is there in infants for sleep-related memory improvements in language-related tasks? How has sleep been found to support memory consolidation and generalisation of newly learnt words in early childhood?

We should note that, in this introduction, we have summarised how some prominent models currently conceptualize the way that sleep affects memory formation and development. These processes are best explained in neurophysiological terms, as sleep-related consolidation processes as currently understood are intimately related to the neural organisation of memory systems, as well as to the electrophysiological state of the sleeping brain. However, most of the papers reviewed here draw on behavioural and linguistic rather than neurophysiological approaches.

The review is structured as follows: First, the methods employed to systematically search and review the literature are outlined. Second, the studies identified are grouped and discussed on the basis of the methods employed and the memory processes investigated. Finally, an integrated discussion of all the studies selected is presented, together with suggestions for future research.

2. Methods

2.1 The approach

The literature review followed the method proposed by Pickering and Byrne (2014). This approach involves explicitly specifying the methods used to identify, review and evaluate the relevant literature, which ensures replicability. Furthermore, the search and review methods that Pickering and Byrne propose make it possible to develop a relatively comprehensive appraisal of the extent to which sleep-associated effects in language learning in infancy have been investigated and, potentially, to identify existing gaps that future research could address. A systematic approach to the review was chosen over a narrative approach for the reasons outlined in Pickering and Byrne (2014). First, narrative reviews are more susceptible to potential bias than systematic reviews. For example, papers may be selected on the basis of the expertise and prominence of particular author(s), or of specific trends in the results they report, such as conformity with the reviewer's goals and expectations. These biases may lead reviewers to limit their searches to specific articles and to overlook other studies that might be equally relevant to their focus of interest. Second, narrative reviews tend to be less clearly organised than systematic reviews, which could make them more subject to the readers' or the authors' subjective interpretations of the results. Clearly presenting each step of the review process instead provides the readers with the opportunity to more objectively appraise both the process itself and its conclusions. Third, as shown in the Results section, the topic investigated is largely multidisciplinary. Limiting the literature search to databases specific to a single discipline and to the reference lists of relevant papers in the field could have led to the omission of relevant studies.

2.2 Inclusion Criteria

2.2.1 Terminology

The first and foremost inclusion criterion concerns the presence of relevant terminology in each study. In order to be included, studies had to contain, within their text, the keywords reported in Table 1. As can be seen in Table 1, the keywords were chosen to specifically address the four main aspects constituting the review question. In addition, the keywords were used to search subject headings, to prevent the research from extending to less pertinent fields while also increasing the chances of including as much of the relevant literature as possible. Initially, the keywords were identified as those used in the most often cited papers. As the search continued, the list of keywords was updated by adding synonyms and new terms. For example, as an increasing number of studies reporting distinct sleep effects on memory were identified, the names of those effects were included in the search (e.g. memory consolidation, abstraction, generalisation, etc.).

Table 1. Keywords used in the searches.

Main topic	Keywords
Sleep	Nap, sleep, sleep-dependent (-associated, -related) memory consolidation
Age	Infant, early childhood, infancy, preschool child, toddler
Memory	Cognitive development, memory consolidation, abstraction, generalisation, memory retention, memory formation, declarative memory
Language	Language development, word learning, language acquisition, vocabulary development

2.2.2 Publication

Only original experimental research was included. This was to ensure that all the papers being considered were primary sources of data and that all would have undergone a peer-review process before being published. Other kinds of publications (e.g. book chapters, monographs, reviews, conference proceedings, etc.) were also consulted in order to better understand and comment on the topics of the review, but were not included among the studies reviewed. Other narrative and systematic reviews on the topic were also consulted, to check that all the relevant literature had been considered to the best extent possible, but these were not included in the analysis. Finally, no limitations were imposed based on year of publication.

2.2.3 Sample

Studies were reviewed if their samples had been selected according to these criteria:

- The children were tested between birth and 3 years of age;
- The children being tested were developing typically, i.e.:
    - Had no known atypical language development (e.g. Selective Language Impairment);
    - Had no known atypical cognitive development (e.g. autism, Down Syndrome);
    - Had no known neurological condition (e.g. epilepsy);
    - Had no known psychological condition (e.g. social anxiety)
- The children being tested had no known specific sleep disturbances (e.g. sleep apnoea, insomnia, etc.).

As the main purpose of this systematic review was to understand the impact of sleep on language learning in early childhood, the primary criterion for inclusion of key literature was the age of the children tested. Studies where children with atypical language development were tested were excluded from the review as this population usually shows unique characteristics with respect to language acquisition which go beyond the scope of this review. Similarly, papers studying children with atypical cognitive development, as well as neurological conditions, were excluded, given the disturbances to both language development and sleep patterns often found in these populations. Studies where the children being tested showed conditions related to sleep were excluded on the same basis, as atypical sleep patterns may impact language and cognitive development in ways that exceed the scope of this investigation.

2.3 Search Strategy and Paper Selection

Four main databases were used for the literature search: PsycINFO, Web of Science, Linguistics & Language Behavior Abstracts (LLBA), and Scopus. These databases are all available online and were chosen as likely to contain papers relating to the fields of Psycholinguistics and Language and Cognitive Development (PsycINFO and LLBA in particular). Scopus and Web of Science were chosen for their more general and interdisciplinary scope, to increase the possibility that publications with a wider scope were also included. The searches were run in December 2019 and again in February 2020. A web-based reference management software tool was used to store and organise the studies included in the review.

Table 2 shows the searches run on each of the databases. Once a first set of keywords was identified through a preliminary consultation of the literature, the search strings were improved as the searches continued. In fact, each time a search was run, the extension and relevance of its results were evaluated and the keywords and search strings modified accordingly, to ensure that a sufficient range of relevant literature was browsed without including a disproportionate amount of literature unrelated to the topics under investigation. At each round of searches, articles were first screened by excluding those that clearly did not match all the inclusion criteria based on the information reported in the abstract or title (e.g. participants were older than 3 years of age or belonged to atypical populations, etc.). Papers were then screened by reading their full text. In addition, the reference list of each paper included in the review was inspected to identify additional relevant literature based on the titles mentioned (i.e. if they contained any of the key words used in the searches).

Table 2. Search strategies.

Databases	**Searches**
PsycINFO	(sleep* OR nap* OR “daytime sleep” OR “nighttime sleep”) AND (child* OR infan* OR toddler* OR “early childhood” OR “preschool child*”) AND (memory AND (“memory consolidation” OR “memory formation” OR “memory” OR “cognitive development” OR “abstraction” OR “generali?ation” OR retention OR retain OR form* OR consolidat*)) AND (language AND (learn* OR acqui* OR develop*))

	OR vocabulary AND (develop* OR acqui* OR develop*) OR "word learning"
LLBA	noft(sleep* OR nap*) AND noft(child* OR infan* OR toddler*) AND noft(memory AND (consolidat* OR form* OR generali?* OR abstract*)) AND noft(language AND (develop* OR acqui* OR learn*)) *noft* = anywhere except full text
Web of Science	TS= (sleep* OR nap*) AND TS= (memory AND (consolid* OR form* OR retain* OR retention OR generali?* OR strength*)) AND TS= (child* OR toddler* OR infan*) AND TS = (language OR vocabulary OR word*) *TS* = topic
Scopus	TITLE-ABS-KEY (sleep* OR nap*) AND TITLE-ABS-KEY (memory AND (consolid* OR form* OR retain* OR retention OR generali?* OR strength*)) AND TITLE-ABS-KEY (child* OR toddler* OR infan*) AND TITLE-ABS-KEY (language OR vocabulary OR word*) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "re")) *TITLE-ABS-KEY* = Document title, abstract and keywords *Ar* = article *Re* = review

2.4 Limitations

The methods described have a number of advantages, as outlined in the first paragraph of this section. However, although using the search strategies typical of systematic reviews can limit the extent to which potentially problematic aspects and biases might occur, it is unlikely that all these factors can be ruled out. For example, potentially relevant publications that are not available on electronic databases could have remained unacknowledged and thus excluded from the review. Moreover, although the vast majority of the published research is in English, articles in other languages (apart from French and Italian, which are accessible to the main researcher) could not be reviewed. Furthermore, as shown in the next section, the number and scope of the papers eventually included in the review did not provide an appropriate form or sufficient data to conduct a meta-analysis, which could lead to a deeper appraisal of the significance of the results and of what is currently known about the topics under investigation. Finally, an unavoidable bias in searching for published articles is the tendency for investigators to publish, and consequently for a reviewer to access, studies with ‘positive’ results only. As a consequence, studies showing no effects of sleep on memory formation in infant language development are less likely to have been published, and would thus be missing from the present review.

3. Results

Database searches returned a total of 308 papers, which were first screened and selected based on their title (leaving 89 articles) and then their abstract (leaving 33 articles). This approach ensured that all of the articles included strictly complied with the inclusion criteria. Twenty-nine articles were identified as relevant after full-text inspection. After excluding 13 duplicates, 16 papers were finally judged to fully meet the inclusion criteria and were included in the review. These studies are listed in Table 3. Given their small number and the variety of methods, ages and variables represented, a meta-analysis of the findings was not attempted.

Table 3. The studies included, ordered by mean age of the youngest children tested (where multiple studies test the same age group, they are ordered by date of publication).

First author	Year	Age of participants (in months)	Title	Journal	Journal field
Dionne	2011	5.39 to 62.64	Associations between sleep-wake consolidation and language development in early childhood: A longitudinal twin study.	Sleep	Sleep and Circadian Science
Friedrich	2017	6.0 to 8.0 (M = 7.2)	The sleeping infant brain anticipates	Current Biology	Biology (all areas)

			development		
Simon	2017	M = 6.21	Sleep confers a benefit for retention of statistical language learning in 6.5 month old infants.	Brain and Language	Language and Cognitive Neuroscience
Horváth	2016	7.73 to 37.83 (initial assessment)	Frequent daytime naps predict vocabulary growth in early childhood	Journal of Child Psychology and Psychiatry	Child and adolescent psychology and psychiatry
Friederich	2015	9.0 to 16.0 (M = 12.20)	Generalization of word meanings during infant sleep.	Nature Communications	Natural sciences
Friedrich	2019	14.0 to 16.0 (M = 15.12)	The reciprocal relation between sleep and memory in infancy: Memory-dependent adjustments of sleep spindles and spindle-dependent improvement of memories.	Developmental Science	Psychology and Developmental Cognitive Neuroscience
Gómez	2006	15.0	Naps promote abstraction in language learning in infants.	Psychological Science	Psychology
Hupbach	2009	M = 15.23	Nap-dependent learning in infants.	Developmental Science	Psychology and Developmental Cognitive Neuroscience
Horváth	2015	16.0	Napping facilitates word learning in early lexical development.	Journal of Sleep Research	Basic and clinical sleep research.
Horváth	2016	16.0	A Daytime Nap Facilitates Generalization of Word Meanings in Young Toddlers.	Sleep	Sleep and Circadian Science
He	2020	25.1 – 29.9 (M = 26.8)	Two-year-olds consolidate verb meanings during a nap.	Cognition	Interdisciplinary research on the study of mind
Munro	2012	29.0 – 36.0 (M = 32.65)	Why word learning is not fast	Frontiers in Psychology	Developmental psychology
Werchan	2014	30.0 – 35.0 (M = 32.94)	Wakefulness (Not Sleep) Promotes Generalization of Word Learning in 2.5-Year-Old Children.	Child Development	Child development
Axelsson	2018	29.8	Napping and toddlers' memory for fast-mapped words	First Language	Child language acquisition
Williams	2014	41.26 – 43.14 (Ms)	Goodnight book: sleep consolidation improves word learning via storybooks.	Frontiers in Psychology	Developmental psychology
Sandoval	2017	35.22 – 41.29 (M = 37.18)	Words to sleep on: Naps facilitate verb	Child Development	Child development

			generalization in habitually and nonhabitually napping preschoolers		

3.1 What is the evidence for sleep-related improvements in infant memory in language-related tasks?

All of the studies reviewed except Werchan & Gómez (2014) identified sleep-associated enhancement of memory for language in infancy and early childhood. These positive results were found across different age ranges, research designs and methods. In addition, different studies focused on sleep-related improvements in different aspects of memory formation and word learning, such as in the consolidation of new word form-meaning associations (e.g. Horváth et al., 2015), generalisation at the level of word meanings (e.g. Friedrich et al., 2015), and the extraction of phonological structure (e.g. Hupbach et al., 2009). Two studies (Dionne et al., 2011 and Horváth et al., 2016) were based not on experimental data but on longitudinal survey data, i.e. the Quebec Newborn Twin Study in Dionne et al. (2011) and the Oxford Communicative Development Inventory (OCDI) in Horváth et al. (2016). The results of the studies will now be presented, organized by the methods employed and the level of language learning investigated.

To study the relationship between sleep and language processes, most studies have employed a cross-sectional design, comparing children’s memory performance after a learning experience or a task, following a delay during which one subgroup had a nap or overnight sleep and the other did not. In such studies, no experimental manipulation of sleep duration or occurrence (such as sleep deprivation) is involved. Instead, the learning and testing phases were scheduled around each child’s typical sleeping times to the greatest extent possible, to avoid participant distress and the influence of potentially confounding variables (e.g., tiredness).

3.1.1 Methods employed

Appendix A summarises the methods used in each study, with respect to how sleep was measured, the materials employed, the learning and testing procedures administered to participants and whether or not a measure of existing vocabulary knowledge was included.

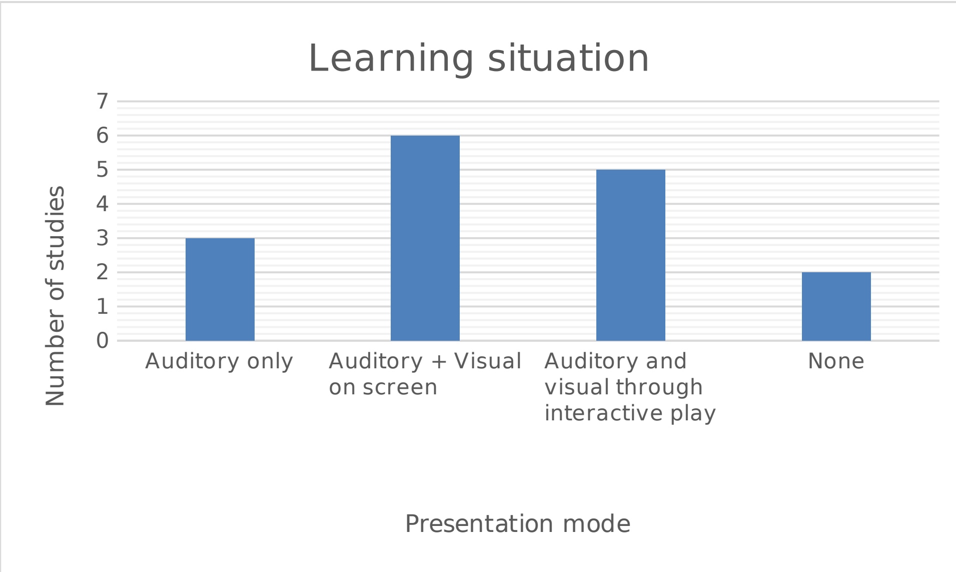
3.1.1.1 Sleep Measures

Sleep was assessed via reports and questionnaires completed by parents in seven out of 16 studies (43.75%). Of these studies, three employed additional physiological measures (i.e. polysomnography or actigraphy); Werchan et al. (2014), He et al. (2020) and the two longitudinal studies by Horváth et al. (2016) and Dionne et al. (2011) did not. Six studies out of 16 (37.5%) measured sleep via actigraphy. In all but one of these studies (i.e. Axelsson et al., 2019) actigraphy was used in conjunction with other measures, as reported in Appendix B. Polysomnography (PSG) was the only sleep measure employed in three out of 16 studies (25%) and, in Simon et al. (2017), it was used together with a parental questionnaire.

3.1.1.2 Learning and Testing Procedures

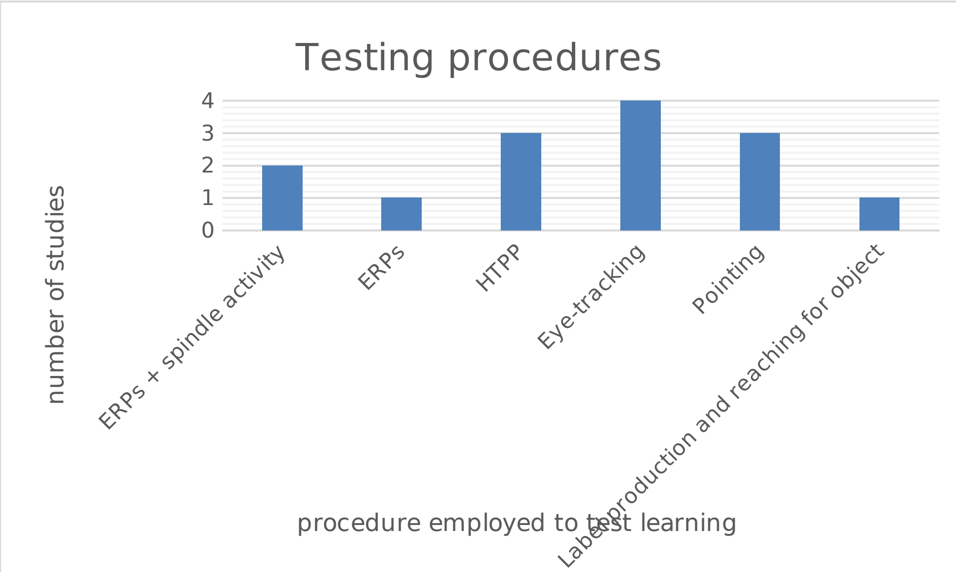
Three studies (Gómez et al., 2006; Hupbach et al., 2009; Simon et al., 2017) targeted word-form learning, using an artificial language task. The artificial language strings included no pauses or prosodic cues that could suggest word boundaries. All three studies employed stimuli that were auditory strings of an artificial language and involved no association of meaning with the word forms. Moreover, they were not intended to attract children's focused attention, as children overheard the stimuli in the background while quietly playing with the experimenter. The strings were presented in auditory form only and contained transitional probability cues between the nonwords that composed them. After having been exposed to them, children were tested for their ability to use such cues to detect either nonadjacent relationships between the nonwords or word boundaries and retain them after a delay. In all other studies reviewed here, the tasks involved learning of a sound-meaning pairing. Of these, six (Friedrich et al., 2017, 2015, 2019; He et al., 2020; Sandoval et al., 2017, Werchan & Gómez, 2014) involved a highly-structured experimental setup in which, in the learning phase, the referent stimuli were presented visually, on a screen, while the corresponding word forms were presented auditorily (either the object label or that of the category the object belongs to). Five other studies (Horváth et al., 2015, 2016; Munro et al., 2012; Axelsson et al., 2018; Williams & Horst, 2014) employed a learning phase embedded in a playful activity, during which children received a controlled number of exposures to the novel words and referents. In three of them, referents were physical objects that children could manipulate while the experimenter labelled them through carrier phrases. In the other two (Axelsson et al., 2018; Williams & Horst, 2014), they were represented in pictures. In Axelsson et al. (2018) referents were represented visually in the form of pictures while the experimenter asked the child to point to them, either praising or correcting the child based on response accuracy. In Williams and Horst (2014) children were exposed to pictures of new referents and their labels through book reading and associated activities. As the two longitudinal studies did not involve an experimental manipulation, they had no learning phase. The learning procedures employed in the studies are summarised in Figure 1.

Figure 1: Learning situation



As for the testing procedures (depicted in Figure 2), three studies (19%) employed electrophysiological measures, 11 studies (69%) employed behavioural measures and two studies (13%) employed surveys. In particular, two of the three studies in which electrophysiology was employed measured both ERPs in response to stimulus presentation and spindle activity during sleep after learning, whereas the third study measured Event Related Potential (ERP) responses to the stimuli only. Amongst the behavioural studies, the measures employed were Head Turn Preference Procedure (HTPP) (three), eye-tracking (four), pointing (three) and label production as well as reaching for the object (one). HTPP was used with the youngest age groups, up to 15 months of age, to test children's orientation to word forms only (i.e. with no new meanings involved). Eye-tracking was used from 16 months of age on. Active child's reaching for the object was employed in studies testing children older than 2. All but one study tested learning using only a receptive, not a production task. In Munro et al. (2012), children's recognition was tested by asking them to perform actions with the object corresponding to the target label pronounced by the experimenter. In addition, they were also asked to imitate and spontaneously produce the label of the target object.

Figure 2: Testing procedures employed



3.1.1.3 Studies that Measured Existing Vocabulary Knowledge

Finally, only four studies (i.e. the two survey-based studies and Horváth et al., 2015, 2016) measured the number of words already known by infants at the time of the study and the association of that measure with sleep-related memory processes in language. Overall, the studies all found a significant positive correlation between vocabulary and sleep-related improvements in memory for new words. The findings of these studies will be presented and discussed in more detail below.

3.1.2 Memory Processes Investigated and Main Results

Details about the design and procedures employed as well as the key findings of the studies reviewed are reported in Appendix B. We found that the studies could be grouped on the basis of the memory and language learning processes investigated. In what follows, a narrative synthesis will present the findings of these studies.

3.1.2.1 Consolidation and abstraction of phonological structure

Three studies (Gómez et al., 2006; Hupbach et al., 2009; Simon et al., 2017) examined sleep-associated abstraction of the underlying distributional sound structure of the language. This ability was held to be fundamental to children's acquisition of grammatical structure as well as for the detection of word boundaries. All three studies found that, after sleep, children were better able to detect the positional dependencies between segments composing strings of an artificial language. In two studies (Gómez et al., 2006; Hupbach et al., 2009), 15-month-olds who did not nap after being familiarized with such strings consistently preferred to listen to the strings they had heard during familiarisation. This was taken as evidence for veridical memory for those strings. In contrast, children who took a nap after hearing the strings, while also remembering the strings from familiarisation, quickly developed a preference for the type of string first presented at test, showing a preference for the familiarised strings only if the first test item was a familiarised one, but a preference for the unfamiliarised strings if the first test item was novel. These findings were interpreted as evidence for the nap group having generalised over the strings following sleep, as they seemed to notice structural dependencies in the first test trial and then to preferentially attend to those same structures for the rest of the test. Importantly, Hupbach et al. (2009) found that infants in the nap group who did not nap soon after familiarisation failed to show such generalisation effects 24 hours later, suggesting that timely naps are necessary for successful abstraction. On the other hand, Simon et al. (2017) investigated sleep-related abstraction of the internal structure of an artificial language in terms of 6.5-month-olds' ability to detect transitional probabilities and use them to identify word boundaries and segment continuous speech into individual words. They found no significant differences between the wake and nap groups' performances. Moreover, nappers looked longer to words than to part-words (i.e. segments made of two halves of words which had occurred sequentially in the familiarisation string) in the first block of testing and vice versa in the second, although these differences were not significant. However, the authors identified a significant interaction of testing block by trial type (i.e. Words or Part-words) in nappers and interpreted this as evidence for the nap group significantly preferring either one type of word or the other, and thus for an ability to distinguish between them. However, looking behaviours for words vs. part-words were not different within the nap group and, perhaps more importantly, between groups. This potentially limits the extent to which sleep, the main variable by which the two groups differed, can be assumed to have influenced children's extraction of words from familiarization strings.

3.1.2.1 Learning of sound-meaning pairs.

3.1.2.1.1 Memory consolidation: veridical memory for new word form-meaning associations.

Another line of research specifically investigates the effects of sleep on the consolidation of word form-meaning associations (Horváth et al., 2015; Williams & Horst, 2014; He et al., 2020; Axelsson et al., 2018). In Horváth et al. (2015), 16-month-olds who napped and children who did not nap after learning recognised the newly learnt associations at similar levels when tested immediately after learning. However, when tested after a delay, those who had napped during the delay performed significantly better than they had done at the immediate test, and also significantly better than the non-nappers, whose performance had decreased compared to immediate test. Moreover, while neither receptive nor expressive vocabulary was associated with non-nappers' performance, expressive (although not receptive) vocabulary was strongly and positively associated with the shift in performance from immediate to post-sleep test in nappers. This was the only experimental study to specifically investigate the relationship between sleep-related consolidation and existing vocabulary knowledge in children under the age of 3, but it showed results similar to those observed with older children (e.g. Henderson et al., 2012).

Evidence for such a strong sleep-related improvement in memory consolidation for words and meanings is also found in children around their second birthday. Specifically, Williams & Horst (2014) exposed toddlers to new words and tested them after a nap and again after seven days. Two groups of toddlers heard the words through the same story being repeated several times: the first had a nap between learning and testing (Group 1), the second remained awake (Group 2). Other two groups of toddlers learned the words through different stories: one group napped between learning and testing (Group 3) while the other remained awake (Group 4). Toddlers who heard the words through the same story being repeated several times (i.e., Groups 1 and 2) learned the words best overall, whether they napped or not. However, of the two groups who learned the words through different stories, Group 3, who napped after learning, caught up and reached comparable levels of memory to those achieved by Group 2, who heard the same story but did not nap. In contrast, children in Group 4, who learned the words from different stories but did not nap, never caught up with the other three groups and had the weakest memory for the new words and their meanings overall. Indeed, when story repetition was controlled for, sleep was the strongest predictor of performance on the final test, conducted seven days after the initial exposure. Horváth et al. (2015) and Williams & Horst (2014) focused on the consolidation of new nouns. He et al. (2020) found similar beneficial sleep effects in the consolidation of verb meanings. In their study, 2-year-olds’ looks to a target scene increased across visits if they napped after learning, whereas those of a wake group decreased, suggesting a role for sleep in word-meaning consolidation in verb learning as well.

Two studies (Axelsson et al., 2018; Munro et al., 2012) tested the impact of sleep on toddlers’ long-term memory for words learned in an active way. In these studies, new words for the target items were presented to the children together with foils; the children were asked to point to the targets without prior learning, and were given feedback on their performance. In Axelsson et al. (2018) 2.5-year-olds who did not nap after learning gradually forgot the new word form-meaning associations learned in this way, as their performance became progressively worse from immediate test, over two intermediate delayed tests and an additional one on the following day. In contrast, the nap group’s performance remained constant over the testing sessions and was significantly higher than that of the wake group after a 4-hour delay and even on the following day. Moreover, the longer the children napped, the higher their retention scores were.

Interestingly, despite not being specifically focused on the impact of sleep on memory for new words, Munro et al. (2012) obtained similar results with 3-year-olds. Children were asked to recognise and imitate new words for novel referents on multiple testing occasions (i.e. immediately after the learning phase, and again after 1 minute, 5 minutes and 4 days). While children’s recognition of the new words and referents improved over time, the accuracy of the elicited imitations decreased significantly from one interval to the next. This suggests that while recognition for words learned may improve following off-line consolidation (as seen in Axelsson et al., 2018), production accuracy does not appear to. However, regression analyses revealed that children with larger extant vocabulary reached higher levels of production accuracy, thus suggesting that the representations may be scaffolded by children’s existing vocabulary.

3.1.2.1.2 Generalisation of form - meaning associations.

Four studies (Friedrich et al., 2015; Horváth et al., 2016; Sandoval et al., 2017; Werchan & Gómez, 2014) looked at sleep-related generalisation of word meanings. In these studies, children were familiarised with novel word forms and their referents and then tested on their ability to generalise the new word form to new but similar-looking referents. Thus, in these studies, the new word forms corresponded to the names of new categories of referents. Friedrich et al.’s (2015) electrophysiological study measured infants’ ERP responses to correctly and incorrectly matched word and referent. In previous studies, an increase in negativity had been identified between 200 and 500ms in left lateral regions of the infant brain in response to correctly paired word-forms and meanings; this was taken to be a word-form priming effect, i.e. a sign of a child’s recognition of the association between a word and its specific meaning (N200-500 effect) (Friederici & Thierry, 2008). In addition, increased negativity in the centro-parietal region 400ms after stimulus presentation was interpreted as a sign of higher-level semantic processing (N400 effect: Friederici & Thierry, 2008). In Friedrich et al. (2015), children were exposed to (i) the same objects repeatedly paired with the same labels, (ii) different but similar objects paired with the same category labels and (iii) objects inconsistently paired with different labels. Those who napped after this phase showed an increased N200-500 effect in response to labels correctly paired with the individual objects they stood for and that the children had seen in the learning phase. This was taken to indicate long-term memory for the specific word-meaning associations they learned. In addition, nappers showed an increased N400 effect for category labels that had been incorrectly paired with a novel object (i.e. an object they had never seen before but that was similar to those they had been exposed to in association with another category label). This effect was taken to indicate an effort to integrate mismatching information with existing knowledge, thus suggesting that the word form-meaning association had been generalised. On the other hand, children who did not nap showed none of these brain responses after a delay, which was taken to mean that they had not retained the meaning of the novel words and also did not generalise those words to novel objects of the same category. Therefore, the fact that only nappers showed a response to incorrect associations of objects and category names suggested to these authors that the nappers had formed a memory representation of that category. The authors also confirmed the positive association found in the literature between spindle activity during post-learning sleep and improvements in memory performance after sleep, as they found a positive correlation of spindle activity with the occurrence of the N400 effect in the nap group.

Three other studies investigated the same memory process from a behavioural point of view. Horváth et al. (2016) observed that only children who napped after being familiarised with new word-object associations were able to recognise new but similar objects as members of the same object category. Perhaps even more importantly, the same children were unable to do so immediately after learning. This sleep effect was also found for the generalisation of verb meanings in Sandoval et al. (2017), who tested 3-year-olds. They found that only children who napped after learning managed to correctly recognise the actions corresponding to the new verbs, which they had learnt the day before, when performed by different actors. This was taken as evidence of generalisation of meanings. Non-nappers, regardless of whether they were habitual nappers or not and despite a full night's sleep between learning and testing, still performed significantly worse than nappers. In contrast, Werchan & Gómez (2014) identified what seemed to be a negative effect of sleep on the generalisation of meanings to new object categories. Children (who were 2.5-year-olds) who remained awake in the 4-h delay after learning performed the most accurately, when hearing the category name that they had learned and asked to point to the correct reference in a four-alternative forced-choice task. Moreover, they even performed better than a third group tested immediately after learning; no statistically significant differences were found between the performances of the nap group and the immediate-test group.

3.1.2.2 Changes of brain responses and sleep parameters observed during or after learning.

Two studies (Friedrich et al., 2019; Friedrich et al., 2017) investigated the effects of sleep on memory formation from a more specifically electrophysiological perspective. In particular, they investigated the associations between brain responses related to the learning processes and brain activity during sleep.

Friedrich et al. (2019) found that in 14- to 16-month-olds the number and density of sleep spindles were significantly larger in the post-learning phase (i.e. after learning previously unknown words) than after a non-learning control condition (i.e. where children were exposed to words they already knew as labels of objects of familiar categories). Most importantly, during testing after sleep, the occurrence of the N200-500 effect in the ERP was observed only after novel-word learning and only in those children with higher spindle density increase, meaning that only these children managed to form object categories and to appropriately associate exemplars to them. In addition, an effect similar to the N400 (i.e. one that occurred more than 400ms after stimulus onset but that was analogous to the N400 in terms of the brain regions involved and the polarity) was visible only in children with higher spindle density and mainly after learning. However, some children already showed a late-N400 effect at encoding during the learning session, although it was not as strong a predictor of post-learning generalisation as spindle density. Friedrich et al. (2017) observed similar effects even in younger infants (i.e., 6 to 8 months of age). They hypothesized that longer NREM sleep (when SWS-related memory consolidation normally occurs) could be positively associated with the formation of semantically-based word-meaning links (taken as evidence of a more linguistically mature infant brain) as compared to perceptually-based associations (i.e. associations between holistic sound representations and specific referents or context-specific categories), which according to the authors are most typical of younger infants. Indeed, in another study, Friedrich and Friederici (2017) repeatedly exposed 3-month-olds to consistent and inconsistent initially novel word-object pairings. Then, as they measured the infants' ERPs in response to correct and incorrect pairings immediately after exposure to the stimuli, they observed that infants showed significantly different responses to correct and incorrect pairings, but did not do so on the next day. The 3-month-olds' responses at immediate test were taken to indicate perceptually-based associations rather than responses to semantically-based word-meaning links. In Friedrich et al. (2017), it was found that no ERP effects occurred in the wake groups or in any of the children after immediate test in response to any kind of category-name-to-object pairings. Moreover, 6-to-8-month-olds in the short-nap group showed a similar ERP effect to that observed by Friedrich and Friederici (2017) in 3-month-olds in response to consistent new, individual object-label pairings. In contrast, 6-to-8-month-olds in the long-nap group showed enhanced negativity to incorrect category name-object pairings; that is, they showed an N400 effect, indicating higher-level semantic processing of the stimuli and, most importantly, generalisation of word meanings at the category level.

3.1.2.3 Longitudinal effects of sleep on language development

Finally, two studies (Dionne et al., 2012; Horváth et al., 2016) investigated the longitudinal relationship between current sleeping patterns and later language development. Dionne et al. (2012) observed the associations between the maturation of more adult-like sleep patterns (i.e. a progressive shift from recurrent naps during the day to more prevalent night-time than daytime sleep), measured at 6, 18 and 30 months, and language skills measured at 18, 30 and 60 months. The authors found that a smaller ratio of daytime over night-time sleep at 6 months was associated with larger lexicons at 18 and 30 months and that at 18 months this ratio predicted vocabulary at 60 months. These findings indicate that individual sleep maturation levels are potentially very strong predictors of current and future linguistic abilities in children. However, Horváth et al. (2016) obtained sharply contrasting results. They found that the number of daytime naps and sleep efficiency measured at initial assessment (i.e. between 7.73 and 37.83 months of age) were significant positive predictors of receptive vocabulary growth but only marginally significant positive predictors of expressive vocabulary growth, both measured between 3 and 6 months after initial assessment. Furthermore, in Horváth et al. (2016) night-time sleep duration at initial assessment was negatively correlated with expressive vocabulary development.

4. Discussion

This review has synthesized the findings of studies that investigated the effects of sleep on several memory processes related to language acquisition in children under the age of 3. The large array of methods employed, combined with the variety of age groups investigated and the small number of studies, limits the extent to which the findings of these studies can be statistically analysed and compared. However, common themes can be identified, making it possible to compare the studies in a narrative fashion. In what follows, these themes are addressed and discussed in relation to the current literature and their potential significance is discussed in greater depth. Moreover, the strengths and limitations of the studies reviewed will be outlined, to suggest possible directions for future research.

This review has shown that the findings as to the beneficial effects of sleep on several processes involved in language learning that are well-established in adults (Tamminen et al., 2010; Feld & Diekelmann, 2015) and school-aged children (e.g. Ashworth, Hill, Karmiloff-Smith & Dimitriou, 2014) can be extended to early childhood and infancy. In fact, in infancy as well as in adulthood, sleep seems to play a direct and active role not only in protecting and preserving newly formed memories but also in actively strengthening and enhancing memory performance and generalisation. Direct evidence for such a role for sleep in infant language learning is provided by the studies investigating changes in brain responses and sleep parameters after learning. One example is the increased spindle activity in infant nap sessions following learning of new words compared to nap sessions following exposure to already known words (Friedrich et al., 2019). As outlined in the introduction, sleep spindles are the electrophysiological phenomenon that drives hippocampal memory reinstatement in the neocortex (James et al., 2017). Therefore, these spindles are crucially associated with the process of integration of newly learned memories, as proposed in the ASC model (Born & Wilhelm, 2012). Accordingly, the association between spindle activity and learning of new material observed in Friedrich et al. (2019) clearly indicates an active role of sleep in memory consolidation. Similarly, the positive association between sleep spindle activity in the post-learning nap and the occurrence of the N400 effect in Friedrich et al. (2015), indicating generalisation of word form-meaning associations, strongly supports an active role for sleep in infant word learning.

Indeed, several studies seem to suggest that, in infant learning of novel words in experimental settings, not only sleep in general but naps in particular may play an influential role. First, where memory performance was measured immediately after learning and compared to that observed after a delay, either including a nap or not (e.g. Horváth et al., 2015; Williams & Horst, 2014; Axelsson et al., 2018; Horváth et al., 2016), improvements were found in the second test only in those children who took a nap during the delay; in contrast, non-nappers' performance decreased. Crucially, though, studies where a third test was included after a 24-hour delay (e.g. Williams & Horst, 2014) found that children who had taken a nap after learning still outperformed children in the wake groups the following day, despite the latter having had a full night of sleep during which processes of memory consolidation could be expected to have occurred. The support of the potential importance of timely naps in new word learning is particularly evident in Williams and Horst (2014). Indeed, even though hearing the words through different stories led to less successful learning in general, children who heard the words in different stories and then napped achieved similar levels of performance to children who heard the words in one repeated story but did not nap.

It has been proposed that one of the potential effects of young children’s longer sleeping times as compared with adults and of their polyphasic sleep (or sleep consisting of multiple bouts of sleep), rich in slow wave activity, could be protection against the fragility of their hippocampal memory traces (Kurdziel, Duclos & Spencer 2013). In fact, it has been observed that the hippocampal circuitry reaches full maturation only by age 3-5 (Mullally & Maguire, 2014), in concert with other brain areas whose maturation might influence that of the hippocampus (Jabès & Nelson, 2015). It could be hypothesized that younger children need to sleep more frequently to allow their low-capacity hippocampus to transfer memory traces to the target neocortical areas, which might themselves be gradually maturing. This seems to be supported by the findings in the longitudinal study of Horváth et al. (2016), in which the number of daytime naps at initial assessment predicted future receptive vocabulary growth, whereas nocturnal sleep duration was negatively associated with subsequent expressive vocabulary. However, Dionne et al. (2011) identified better vocabulary in children with monophasic rather than polyphasic sleep. The contradictory findings in these studies make it clear that there is room for future investigations to understand the actual impact of naps and the progressive maturation of adult-like sleep patterns on language acquisition in early childhood. However, it should be considered that, in Horváth et al. (2016), age at first assessment (when sleep data were also collected) varied considerably across children (i.e. from 7.73 to 37.83 months of age). As the relationships between daytime naps and nocturnal sleep seem to vary considerably across early development, collecting sleep data from a sample more restricted in age range could provide a more reliable estimate of the impact of sleep maturation in new word learning. Furthermore, it should be borne in mind that, in both of those longitudinal studies, sleep data were collected via parental report only. The exclusive use of parental reports to assess children’s sleep patterns is problematic, as parents cannot always be fully aware of every aspect of their child’s sleeping behaviour (e.g., of their number of night awakenings).

The two longitudinal studies just mentioned were among the small number of studies including measures of vocabulary as either a dependent or independent variable. While vocabulary (expressive and/or receptive) was the variable on which the effects of sleep were measured in these studies, Horváth et al. (2015) investigated the role of existing vocabulary knowledge in sleep-associated consolidation of newly learned words. The finding that stands out the most clearly is that recognition of the newly learned words was higher in those children with larger expressive – but not receptive – vocabularies. This finding suggests that existing vocabulary knowledge may play an important role in the long-term acquisition of new words. Indeed, the integration of new knowledge in the neocortical system in adults has been shown to be prior-knowledge dependent, in the sense that new information that is more consistent with existing knowledge is likely to be consolidated more rapidly (e.g., McClelland, 2013). This may be true for children and infants as well. In particular, as performance improvement after sleep was found to be associated with expressive but not receptive vocabulary, the specific (independent) influences of expressive and receptive vocabulary are worth investigating further. Interestingly, Henderson, Devine, Weighall & Gaskell (2015) observed a similar positive association between expressive vocabulary and the degree of sleep-related enhancement of consolidation of newly learned words, measured in terms of lexical competition, in school-aged children (as explained below). However, as systematic investigations about the unique role of expressive vs. receptive vocabulary in new word learning remain scarce, it would be premature to attempt to draw conclusions as to the specific role of either kind of vocabulary knowledge in word learning.

As mentioned at the beginning of this section, this review has presented the current evidence about the effects of sleep on several processes involved in new word learning in infancy. These processes were identified as the learning of sound-meaning pairs (encompassing the retention of specific word form-meaning associations and their generalisation to new but similar referents) and the consolidation and abstraction of the sound structure of speech. These processes will now each be addressed and discussed in turn.

As regards to the consolidation and abstraction of phonological structure from artificial language strings, studies investigating these effects have identified significant changes in the children's recognition of the underlying structure of an artificial language before and after the delay and depending on whether the delay included a nap or not. In particular, Gómez et al. (2006) and Hupbach et al. (2009) focused on the children's ability to abstract the distributional structure of artificial language strings, seeing that as evidence of their ability to abstract and generalise the grammatical structure of their language. In these studies, nappers' selective looking preference based on the first test-trial type was taken as evidence for their ability to generalise grammatical structure, as compared to the wake group, who preferentially listened to strings presented at familiarisation and, therefore, was seen as showing veridical memory only. However, despite showing different preferences, both the nap and the wake group showed an ability to distinguish between the familiarised and non-familiarised lists. It should be noted that neither novelty preference (as observed in some of the children in the nap group) nor familiarity preference (as in the wake group) actually involves any form of generalisation to unseen items. Therefore, while these two responses may be taken as evidence for the effect of sleep on preference behaviour based on veridical memory for the strings, the extent to which they can be taken as evidence of generalisation is debatable.

Consolidation of the actual word form-meaning associations encountered in the learning phase, observed in veridical memory for novel words, can be considered to be the behavioural effect of the memory consolidation and synaptic consolidation process described in the introduction (Born et al., 2006). When the consolidation of new word form-meaning pairs is investigated, this consolidation is usually assessed by measuring delayed recognition or recall of the word-referent mapping presented in the learning phase (Henderson, Weighall & Gaskell, 2013). As observed in the studies reviewed, an offline period of consolidation containing sleep appears to protect memory for the newly learned sound-meaning pairs from decay, as in Axelsson et al. (2018), where nappers' levels of recognition remained constant over multiple testing occasions while wake group's recognition levels steadily decreased. In addition, as evidenced in Horváth et al. (2015), veridical memory can even be enhanced following sleep, compared to its levels at test immediately following learning, before sleep. But being able to generalise these associations to different but similar referents is also important, as it leads to the construction of categories and to the ability to recognise new objects and events on the basis of differences and similarities with previously encountered instances. This gives rise to knowledge that transcends memory for single items.

As outlined in the introduction, memory generalisation relies on multi-item generalisation, entailing the repeated reactivation of a memory trace in the hippocampus during sleep, which strengthens the representation's invariant characteristics while weakening inconsistent details. In fact, as can be seen in most of the studies reviewed, the generalisation of word form-meaning associations seems to be absent at immediate test (as in Horváth et al., 2016) and to emerge only after a delay including sleep (apart from Friedrich et al., 2019, where some infants showed an N400 effect even before sleep). Indeed, generalisation can be considered to result from system consolidation, i.e. the process by which new neural representations are integrated and reorganised into existing neural networks, which may take days or even weeks to complete and which, crucially, relies on sleep (Born, Rasch & Gais, 2006).

A parallel phenomenon to multi-item generalisation that may result from similar system consolidation processes is memory integration, i.e. the assimilation of encoded memory representations into existing networks and schemas (Stickgold & Walker, 2013). In new word learning, lexical integration is also known as lexicalization (Dumay & Gaskell, 2007). This phenomenon consists in the complete integration of a newly learnt word into the existing lexical network, which thereby results in that word undergoing similar processes to other established lexical entries. The acid test employed by Gaskell and his colleagues to assess lexical integration of newly learnt words is to measure the extent to which they compete with other similar-sounding lexical entries. The underlying assumption is that every word in the lexicon is a node in a network; activation of one node can spread to its neighbours. Lexicalization has been investigated in adults (e.g. Dumay & Gaskell, 2007; Henderson, et al., 2015; Leach & Samuel, 2007) and school-aged children (e.g.

Henderson, Weighall, Brown & Gaskell, 2012). However, up to now, lexical integration in preschool children has seldom if ever been addressed. This is a potentially important issue for future research, given the impressively rapid growth of the lexical network observed early in development.

Similarly, these generalisation processes may underlie generalisation of word forms alone, e.g. across different voices, as in the studies by Houston and Jusczyk (2000; 2003). Indeed, recognition of a word form when uttered in different voices entails the extraction of invariant acoustic characteristics from a pool of several exemplars of that word, which is thus a form of multi-item generalisation. Houston and Jusczyk's studies did not control for the role of sleep in infants' ability to generalise the phonological form of newly learnt words across different speakers. However, they investigated the long-term effects of indexical properties of speech on new word memory representations by testing children's recognition of the new words either immediately or one day after familiarisation. Therefore, although their studies did not specifically investigate the processes underlying consolidation and generalisation of word forms, their findings may well indicate a potentially interesting direction for future research. Indeed, further studies could seek to understand whether infant sleep plays a role in the generalisation of word forms as it does for word meanings, thus helping to paint a more comprehensive picture of the development of generalisation in language learning.

Among the papers reviewed, only Werchan & Gómez (2014) failed to find beneficial effects of sleep on new word learning. Actually, in their study, the nap group seemed to perform worse than the wake group in the generalisation of meanings to new object categories. As mentioned in the last section, children who did not nap performed better at the delayed test than either the nap group or the control group tested immediately after exposure; there were no significant differences between the latter two groups' performance. To explain these findings, the authors hypothesize that the forgetting of details during wakefulness promotes retention of consistent, recurring features that are the basis of meaning generalisation. In their view, this could be particularly frequent in young children, who may be more prone to encode irrelevant along with relevant information. However, such an interpretation seems to contradict what has been consistently evidenced, across cognitive domains and in children of various ages, about the role of sleep in memory consolidation and generalisation. Rather, the immediate-test group’s poor performance, apart from potentially resulting from fatigue directly after participating in the task, might have emerged from possible differences between the groups’ learning capacities, which might have been present already at the outset. Including an immediate test for all the children would have made it possible to rule out any such differences between groups. However, few studies have included an immediate test for all groups. Future research should consider including a pre-exposure or pre-learning test as well as multiple tests over time (e.g. after a nap and after a night’s sleep), as in Williams and Horst (2014). Although each testing experience can also serve as a reminder or an opportunity for learning, including multiple tests would provide a better idea of the evolution of memory representations and of their stability over time.

To conclude, a few methodological suggestions can be advanced on the basis of the findings of the studies reviewed. First, with regards to how sleep patterns are measured, almost all studies combined multiple measures, both physiological and parental-report-based. However, some (e.g. He et al., 2020; Dionne et al., 2011) used only parental reports of the children’s sleeping patterns. Combining physiological measures and parental reports is advisable, in the interest of providing reliable results.

Another methodological point is that, in most of these studies, infants received relatively artificial, passive and massed exposure to the stimuli. In fact, only five studies employed exposure methods that gave an active role to the child learners (e.g. by allowing them to manipulate objects while the experimenter labelled them). Future research would do well to adopt learning procedures that are more similar to the way in which children are naturally exposed to objects and their names, to increase ecological validity and to more reliably assess the mechanisms that underlie natural language acquisition. Indeed, active motoric experience with language, objects and events, and the proprioceptive feedback the child receives from their own actions, are basic to cognitive development (Thelen & Smith, 1994).

One last methodological remark concerns those studies employing exclusively electrophysiological measures. In Friedrich et al. (2019), the emergence of the N400 effect in some of the children tested was unclear (i.e. the effect did not always appear as an N400 and some children already showed such an effect at training). Including a behavioural test in combination with ERP measures would help to clarify cases like that. For example, if children show an ability to generalise (e.g. in a test like that used in Horvath et al., 2016), and their accuracy in the test correlates positively with the N400 occurrence, this would provide more convincing evidence of successful generalisation than interpretation of the ERP results alone.

Conclusions

It is widely known and accepted that sleep has beneficial effects on adult memory evolution and lexical learning. Research investigating and confirming a similar positive role for sleep in memory formation and new word learning in childhood is also increasing. The aim of the present review was to quantify and assess the breadth and depth of current literature investigating such effects of sleep in early childhood, i.e. before age 3. Overall, the findings of this review suggest that sleep plays a key positive and direct role in new word learning in early childhood, over a wide age range (from around 6 months to 3 years of age) and based on the memory mechanisms involved. The insights gained from the results of the studies reviewed and their discussion is intended to lay the groundwork for future experimental investigations in this field, which is promising but still emerging. Indeed, considerably more work is needed to understand the effects of sleep on infant lexical development and its interactions with other factors, such as existing vocabulary knowledge. Future research along these lines may have important implications, given young children’s unique sleep behaviours as well as the fast rate at which they learn language and increase their vocabulary.

References

Ashworth, A., Hill, C. M., Karmiloff-Smith, A. & Dimitriou, D. (2014). Sleep Enhances Memory Consolidation in Children. *Journal of Sleep Research,* 23 (3), 302–8. https://doi.org/10.1111/jsr.12119

Axelsson, E. L., Williams, S. E. & Horst, J. S. (2016). The Effect of Sleep on Children's Word Retention and Generalization. *Frontiers in Psychology,* 7, 1192. https://doi.org/10.3389/fpsyg.2016.01192

Axelsson, E. L., Swinton, J., Winiger, A. I. & Horst, J. S. (2018). Napping and Toddlers' Memory for Fast-Mapped Words. *First Language,* 38 (6), 582–95. https://doi.org/10.1177/0142723718785490

Born, J., Rasch, B. & Gais, S. (2006). Sleep to Remember. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry,* 12 (5), 410–24. https://doi.org/10.1177/1073858406292647

Born, J., & Wilhelm, I. (2012). System Consolidation of Memory during Sleep. *Psychological Research,* 76 (2), 192–203. https://doi.org/10.1007/s00426-011-0335-6

Davis, M. H. & Gaskell, M. G. (2009). A Complementary Systems Account of Word Learning: Neural and Behavioural Evidence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences,* 364(1536), 3773–3800. https://doi.org/10.1098/rstb.2009.0111

Diekelmann, S. & Born, J. (2010). The Memory Function of Sleep. *Nature Reviews. Neuroscience* 11 (2), 114–26. https://doi.org/10.1038/nrn2762

Dionne, G., Touchette, E., Forget-Dubois, N., Petit, D., Tremblay, R. E., Montplaisir, J. Y. & Boivin, M. (2011). Associations between Sleep-Wake Consolidation and Language Development in Early Childhood: A Longitudinal Twin Study. *Sleep,* 34 (8), 987–95. https://doi.org/10.5665/SLEEP.1148

Dumay, N. & Gaskell, M. G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. ***Psychological Science,* 18 (1), 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x**

Feld, G. B. & Diekelmann, S. (2015). Sleep Smart-Optimizing Sleep for Declarative Learning and Memory. *Frontiers in Psychology,* 6 (May), 622. https://doi.org/10.3389/fpsyg.2015.00622

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*(4), 409–438.French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. Applied Psycholinguistics, 29, 463-487. https://doi.org/10.1016/j.wocn.2005.08.002

Friederici, A., & Thierry, G. (Eds.). (2008). Early Language Development: Bridging brain and behaviour. Trends in Language Acquisition Research (TiLAR), Vol. 5, Amsterdam: John Benjamins.

Friedrich, M., Wilhelm, I., Born, J. & Friederici, A. D. (2015). Generalization of Word Meanings during Infant Sleep. *Nature Communications,* 6 (January), 6004. https://doi.org/10.1038/ncomms7004

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., Mölle, M., Born, J. & Friederici, A. D. (2017). The Sleeping Infant Brain Anticipates Development. *Current Biology, CB* 27 (15), 2374–80.e3. https://doi.org/10.1016/j.cub.2017.06.070

Friedrich, M., Mölle, M., Friederici, A.D. & Born, J. (2019). The Reciprocal Relation between Sleep and Memory in Infancy: Memory-Dependent Adjustment of Sleep Spindles and Spindle-Dependent Improvement of Memories. *Developmental Science,* 22 (2), e12743. https://doi.org/10.1111/desc.12743

Gais, S., Mölle, M., Helms, K. & Born, J. (2002). Learning-Dependent Increases in Sleep Spindle Density. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience,* 22(15), 6830–34. https://doi.org/10.1523/JNEUROSCI.22-15-06830.2002

Galland, B. C., Taylor, B. J., Elder, D. E. & Herbison, P. (2012). Normal Sleep Patterns in Infants and Children: A Systematic Review of Observational Studies. *Sleep Medicine Reviews,* 16 (3), 213–22. https://doi.org/10.1016/j.smrv.2011.06.001

Gómez, R. L., Bootzin, R. R. & Nadel, L. (2006). Naps Promote Abstraction in Language-Learning Infants. ***Psychological Science,* 17 (8), 670–74. https://doi.org/10.1111/j.1467-9280.2006.01764.x**

He, A. X., Huang, S., Waxman, S. & Arunachalam, S. (2020). Two-Year-Olds Consolidate Verb Meanings during a Nap. *Cognition,* 198 (January), 104205. https://doi.org/10.1016/j.cognition.2020.104205

Henderson, L. M., Weighall, A. R., Brown, H. & Gaskell, M. G. (2012). Consolidation of Vocabulary Is Associated with Sleep in Children. *Developmental Science,* 15 (5), 674–87. https://doi.org/10.1111/j.1467-7687.2012.01172.x

Henderson, L., Weighall, A. & Gaskell, G. (2013). Learning New Vocabulary during Childhood: Effects of Semantic Training on Lexical Consolidation and Integration. *Journal of Experimental Child Psychology,* 116 (3), 572–92. https://doi.org/10.1016/j.jecp.2013.07.004

Henderson, L., Devine, K., Weighall, A. & Gaskell, G. (2015). When the Daffodat Flew to the Intergalactic Zoo: Off-Line Consolidation Is Critical for Word Learning from Stories. *Developmental Psychology,* 51 (3), 406–17. https://doi.org/10.1037/a0038786

Horváth, K., Myers, K., Foster, R. & Plunkett, K. (2015). Napping Facilitates Word Learning in Early Lexical Development. *Journal of Sleep Research,* 24 (5), 503–9. https://doi.org/10.1111/jsr.12306

Horváth, K., Liu, S. & Plunkett, K. (2016). A Daytime Nap Facilitates Generalization of Word Meanings in Young Toddlers. *Sleep,* 39 (1), 203–7. https://doi.org/10.5665/sleep.5348

Horváth, K. & Plunkett, K. (2016). Frequent Daytime Naps Predict Vocabulary Growth in Early Childhood. *Journal of Child Psychology and Psychiatry, and Allied Disciplines,* 57 (9), 1008–17. https://doi.org/10.1111/jcpp.12583

Houston, D. M. & Jusczyk, P. W. (2000). The Role of Talker-Specific Information in Word Segmentation by Infants. *Journal of Experimental Psychology. Human Perception and Performance,* 26 (5), 1570–82. https://doi.org/10.1037/0096-1523.26.5.1570

Houston, D. M. & Jusczyk, P. W. (2003). Infants' Long-Term Memory for the Sound Patterns of Words and Voices. *Journal of Experimental Psychology. Human Perception and Performance,* 29 (6), 1143–54. https://doi.org/10.1037/0096-1523.29.6.1143

Hupbach, A., Gomez, R. L., Bootzin, R. R. & Nadel, L. (2009). Nap-Dependent Learning in Infants. *Developmental Science,* 12 (6), 1007–12. https://doi.org/10.1111/j.1467-7687.2009.00837.x

Jabès, A., & Nelson, C. A. (2015). 20 Years after "The ontogeny of human memory: A cognitive neuroscience perspective," where are we? *International Journal of Behavioral Development*, 39(4), 293–303. https://doi.org/10.1177/0165025415575766

James, E., Gaskell, M. G., Weighall, A. & Henderson, L. (2017). Consolidation of Vocabulary during Sleep: The Rich Get Richer? *Neuroscience and Biobehavioral Reviews* 77, 1–13. https://doi.org/10.1016/j.neubiorev.2017.01.054

Konrad, C., Seehagen, S., Schneider, S. & Herbert, J. S. (2016). Naps Promote Flexible Memory Retrieval in 12-Month-Old Infants. *Developmental Psychobiology,* 58 (7), 866–74. https://doi.org/10.1002/dev.21431

Kurdziel, L., Duclos, K. & Spencer, R. M. C. (2013). Sleep Spindles in Midday Naps Enhance Learning in Preschool Children. *Proceedings of the National Academy of Sciences of the United States of America,* 110 (43), 17267–72. https://doi.org/10.1073/pnas.1306418110

Leach, L. & Samuel, A. G. (2007). Lexical Configuration and Lexical Engagement: When Adults Learn New Words. ***Cognitive Psychology,* 55 (4), 306–53. https://doi.org/10.1016/j.cogpsych.2007.01.001**

Lewis, P. A. & Durrant, S. J. (2011). Overlapping Memory Replay during Sleep Builds Cognitive Schemata. *Trends in Cognitive Sciences,* 15 (8), 343–51. https://doi.org/10.1016/j.tics.2011.06.004

McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. (1995). Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory. *Psychological Review,* 102 (3), 419–57. https://doi.org/10.1037/0033-295X.102.3.419

McClelland, J. L. (2013). Incorporating Rapid Neocortical Learning of New Schema-Consistent Information into Complementary Learning Systems Theory. *Journal of Experimental Psychology. General,* 142 (4), 1190–1210. https://doi.org/10.1037/a0033812

Mullally, S. L. & Maguire, E. A. (2014). Learning to Remember: The Early Ontogeny of Episodic Memory. *Developmental Cognitive Neuroscience,* 9 (July), 12–29. https://doi.org/10.1016/j.dcn.2013.12.006

Munro, N., Baker, E., McGregor, K., Docking, K. & Arculi, J. (2012). Why Word Learning Is Not Fast. *Frontiers in Psychology,* 3 (February), 41. https://doi.org/10.3389/fpsyg.2012.00041

Pickering, C. & Byrne, J. (2014). The Benefits of Publishing Systematic Quantitative Literature Reviews for PhD Candidates and Other Early-Career Researchers. *Higher Education Research & Development,* 33 (3), 534–48. https://doi.org/10.1080/07294360.2013.841651

Rasch, B. & Born, J. (2013). About Sleep's Role in Memory. *Physiological Reviews,* 93 (2), 681–766. https://doi.org/10.1152/physrev.00032.2012

Sandoval, M., Leclerc, J. A. Gómez, R. L. (2017). Words to Sleep On: Naps Facilitate Verb Generalization in Habitually and Nonhabitually Napping

Preschoolers. *Child Development,* 88 (5), 1615–28. https://doi.org/10.1111/cdev.12723

Seehagen, S., Konrad, C., Herbert, J. S. & Schneider, S. (2015). Timely Sleep Facilitates Declarative Memory Consolidation in Infants. *Proceedings of the National Academy of Sciences of the United States of America*, 112 (5), 1625–29. https://doi.org/10.1073/pnas.1414000112

Seehagen, S., Zmyj, N. & Herbert, J. S. (2019). Remembering the Context of Internal States: The Role of Sleep for Infant Memory. *Child Development Perspectives,* 13(2), 110-15. https://doi.org/10.1111/cdep.12321

Simon, K. N. S., Werchan, D., Goldstein, M. R., Sweeney, L., Bootzin, R. R., Nadel, L. & Gómez, R. L. (2017). Sleep Confers a Benefit for Retention of Statistical Language Learning in 6.5month Old Infants. *Brain and Language,* 167, 3–12. https://doi.org/10.1016/j.bandl.2016.05.002

Stickgold, R. & Walker, M. P. (2013). Sleep-Dependent Memory Triage: Evolving Generalization through Selective Processing. *Nature Neuroscience,* 16 (2), 139–45. https://doi.org/10.1038/nn.3303

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J. & Gaskell, M. G. (2010). Sleep Spindle Activity Is Associated with the Integration of New Memories and Existing Knowledge. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience,* 30 (43), 14356–60. https://doi.org/10.1523/JNEUROSCI.3028-10.2010

Thelen, E. and Smith, L.B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*, MIT Press.

Werchan, D. M. & Gómez, R. L. (2014). Wakefulness (not Sleep) Promotes Generalization of Word Learning in 2.5-Year-Old Children. *Child Development,* 85 (2), 429–36. https://doi.org/10.1111/cdev.12149

Williams, S.E. & Horst, J. S. (2014). Goodnight Book: Sleep Consolidation Improves Word Learning via Story Books. *Frontiers in Psychology,* 5 (March), 184. https://doi.org/10.3389/fpsyg.2014.00184

Wojcik, E. H. (2013). Remembering new words: integrating early memory development into word learning. *Frontiers in Psychology,* 4 (April), 151. https://doi.org/10.3389/fpsyg.2013.00151

Appendix A: Methodological overview of the studies.

	Sleep measures				Materials				Training			Testing						Vo ca b.
												Neuro		Behav.				
Studies by 1st	**Actigraphy**	**PSG**	**Parents'**	**Sleep**	**Artificial**	**Category-word**	**New word-**	**New word-verb**	**Passive**	**Screen**	**Activity (play,**	**Sleep spindles**	**ERP response**	**HTPP**	**Eye-tracking**	**Pointing**	**Production**	
Dionne			x															x
Friedri		x				x				x		x	x					
Simon		x		x	x				x					x				
Horvát			x															x
Friedri		x				x				x			x					
Friedri		x				x				x		x	x					
Gómez	x		x		x				x					x				
Hupba	x		x		x				x					x				

	Sando	Willia	Axelss	Werch	Munro	He	Horvát	Horvát
	x		x				x	x
	x			x		x		
							x	x
				x				
		x	x		x			x
	x					x	x	
	x			x		x		
		x	x		x		x	x
						x	x	x
	x	x	x	x	x			
					x			
					x		x	x

Appendix B: Details about the design, procedures and results of the studies reviewed.

Study	N. of participants	Age range	Design	Stimuli	Procedure	Main findings
Gómez et al. (2006)	24	15.23 months	Cross-sectional (nap group vs. wake group + nap-control group)	Artificial language (no prosodic cues suggesting word boundaries; three-nonword segments, where the first nonword predicts the third, and the medial nonword varies a lot. For the nap-controls, the medial nonword varied less).	Children were exposed to the strings while engaged quietly in other activities. Then, they either napped or remained awake. Then, their recognition was tested via HTPP, when they were presented with familiar strings vs. strings with the same words but in a different order from familiarization.	The wake group showed preference for familiar strings, i.e. the same played during familiarization. The nap group's looking preference was determined by first test trial.
Hupbach et al. (2009)	24	15.30 months	As in Gómez et al. (2006)	As in Gómez et al. (2006)	As in Gómez et al. (2006), but children were tested both after 4h delay (either inclusive of nap or not) and after 24h from exposure to the strings.	Same as Gómez et al. (2006). In addition, among the nappers, those who did not nap shortly after exposure to the strings did not show preference determined by first trial 24h later.
Simon et al. (2017)	37	6.4 – 7.2 months	Cross-sectional (nap group vs. wake group)	Artificial language made from monotone syllables, no cue suggesting word boundaries.	Children were exposed to the strings while engaged quietly in other activities. Then, children either napped or	No statistically significant difference was found between the two groups. Nap group's preference (non-significant) was found for part-

					remained awake. Then, their recognition was tested via HTPP, where they were presented with words vs. part-words (i.e. formed by two adjacent halves of two contiguous words).	words over words in Block 1 (*p* = .068) and vice-versa in Block 2 (*p* = .092). A significant interaction of Block (i.e. 1 or 2) by Trial type (i.e. Words or Part-words) was found in nappers.
Friedrich et al. (2015)	90	9-16 months (mean: 12.65)	Cross-sectional (nap group vs. wake group)	Auditory: disyllabic pseudowords as labels of new object category. Visual: pictures of pseudo-objects belonging to the categories.	Children had on-screen exposure to the stimuli. Then, after a delay where only half of the children had a nap, their ERPs at test were recorded, in two conditions: incorrect vs. correct category pairing conditions. In addition, their sleep spindles activity was recorded.	The wake group showed no N200-500 nor N400 effect (i.e. they did not retain the links between the pseudowords and the object they stood for and were unable to generalise it to new referents) In the nap group, both N200-500 to correct pairings and N400 effect to incorrect pairings were found. The nap group's N400 effect was positively associated with sleep spindles activity during nap after learning.
Horváth et al. (2016)	28	16 months	Cross-sectional (nap vs. wake group)	Two new category labels (pseudowords), four new objects belonging to each category (two original	Learning: interactive play, where the experimenter introduced the two new toys and their names (original	Only children who napped after learning showed a looking preference for the correct object-

				objects and two generalisation objects).	objects) and two similar objects, only labelled with “it” or “this” (generalisation objects)). Then, children had on-screen training and immediate test, followed by 1.5 delay (where half of the children had a daytime nap in the lab and the other half had a walk into town). After the delay children were tested again on their looking to correct vs. incorrect assignment of the generalisation objects to their category labels.	category assignments; non-nappers showed no preference. Moreover, the generalisation effect was visible only when tested after the retention interval inclusive of sleep, not at immediate test. The two groups’ performance was the same at immediate test.
Sandoval et al., (2017)	39	3;0	Cross-sectional (habitual nappers – nap group; habitual nappers-wake group; nonhabitual nappers-nap group; nonhabitual nappers-wake group).	Two novel pseudowords meaning two distinct whole-body actions (video recorded).	On-screen exposure to the new actions and their labels. Children were asked to do their best to learn the new actions. The two nap groups had a nap after exposure. All four groups were tested on the following day on the recognition of the same actions performed by different actors.	Only the two nap groups correctly recognised the actions corresponding to the newly learned verbs, even when performed by different actors, thus showing evidence of generalisation. Non-nappers, regardless of their napping habits and despite the full night of sleep between learning and

						testing, still performed significantly worse than nappers.
Werchan & Gómez (2014)	30	2;5	Cross-sectional (nap group; wake group, immediate test group)	Pictures of three different but similar object exemplars belonging to each category. Each category was labelled with a different pseudoword.	At learning, children were presented with the three objects in each category labelled with carrier phrases (e.g. Look at the *dax!*) one after the other and a distractor object. The test was a forced four-choice alternative recognition test, between a new instance of the object category, a novel object, the distractor and a familiar toy.	Significantly better performance in the recognition test was found in the no-nap than the nap group and the immediate-test group. No statistically significant differences were found between the nap and immediate-test group's performance.
Horváth et al. (2015)	38	16 months	Cross-sectional (nap group vs. wake group)	Two new pseudoword-object pairings	Learning: interactive play where the experimenter introduced two new objects one at a time, each labelled nine times with carrier sentences. This was followed by on-screen training (one block of two familiar words (presented individually either L or R) and one block of two new words (presented individually either L or R)). Then,	The two groups' performance was not significantly different at immediate test. At delayed test, nappers' performance had significantly increased whereas that of non-nappers decreased compared to immediate test, and the former significantly outperformed the latter. Neither receptive nor expressive

					children had an immediate test (IPL), where a block of two familiar words and a block of two new words were presented. After a delay where half of the children had a nap and half remained awake, testing was repeated with the same procedure.	vocabulary was associated with performance in the no-nap group (*r* = -0.34, *p* = 0.198 and *r* = -0.048, *p* = 0.859, respectively). Expressive vocabulary (but not receptive) was strongly and positively associated with the shift in performance from immediate to post-sleep test (*r* = 0.668, *p* = 0.013).
Williams & Horst (2014)	48	3;0	Cross-sectional (four groups: same story-nap group; different stories-nap group; same story-wake group; different stories-wake group)	Two new pseudoword-object pairs.	The new words were embedded in story books. Half of the children were read the same story three times; the other half were read three different books. After being read the book(s), children's initial recognition of the new words was tested via a pointing game. Then, the nap groups took their nap while the no-nap groups played. Recognition was tested again after 2.5h, 24h and 7 days from initial exposure.	The same story-nap group performed the best. The same story-wake group followed them. Different stories-nap group performed just as well as same story-wake group after a night's sleep and continued to perform just as well as same story-wake group. Different stories-wake group performed the worst in all retention tests. Children performed significantly better after 24h than after 2.5h or

						immediate test. Also performed better after 7 days than immediate test. After 2.5h, story repetition predicted 23% of the variance; sleep and story repetition together predicted 50%. After 24h, sleep and story repetition together predicted 39%. After 7 days, nap became the strongest predictor when story repetition was controlled for.
Xiaoxue He et al. (2020)	42	2;3 years	Cross-sectional (nap vs. wake group).	Videos of actions performed by actors (targets were always causative actions)	First, children were provided only with linguistic context anticipating the target actions (i.e. videos of dialogues where one speaker talks about the target action, e.g. “the man is going to *moop* the lady”). Four hours later, they are shown the target and a distractor action and are prompted to look at the target action	In the nap group, the attention devoted to the target scene increased from visit 1 to 2. In the wake group, it decreased.

					(e.g. a voice saying “do you see mooping? Find mooping!”).	
Friedrich et al. (2019)	30	14 to 16 months	Within-subject design (three lab sessions, each on a different day)	Eight exemplars of eight different similarity-based categories (nonwords as category labels). At test, each category had four more exemplars. Nonlearning control session: eight exemplars of eight known categories.	On day 1, learning: exposure to new similarity-based categories (nonwords as category labels), consistently paired, and to object-category inconsistent pairings. On day 2, children had memory test, i.e. recording of their ERP responses to new exemplars which were either correctly or incorrectly paired to their category. On day 3 (one week later), non-learning control session: children had the same test procedure with exemplars of known categories (dog, ball, etc.). After learning and control sessions, all infants napped.	In after-learning nap (day 1), infants had longer stage 2 NREM and more frequent and more intense central parietal spindles compared to the nap following the non-learning session. After the learning session, only the group with bigger spindle density growth showed the word form priming effect. Only the group with high density showed parietal memory effects, whose polarity and distribution resemble to N400. Neither N200-500 nor N400 effects were present in training session. However, some children with weak spindle activity did show N400 in 2^{nd} half of training, while children with strong

						spindles did not. However, immediate generalisation did not predict post sleep generalisation as strongly as spindle activity.
Friedrich et al. (2017)	107	6 to 8 months	Cross-sectional (short-nap, long-nap, wake group). Prior to visit, assigned either to wake or nap based on usual nap times. Then, nappers were classified according to their spontaneous sleep duration. As the wake group had a shorter delay, because not able to say awake for that long, delay length was a covariate.	As in Friedrich et al., 2019	Learning: stimuli were presented in inconsistent pairing condition (i.e., every word paired to every object once, to teach word form but to avoid form-meaning pairing) and in consistent pairing condition (i.e. category words were paired with eight exemplars each). At test, ERP responses to correct- and incorrect category-exemplar pairings (with four new exemplars) were recorded.	Immediate test and wake group showed no effect. Short nap infants showed brain responses of 3-month-olds (perceptual-associative memory), i.e. late negativity. Long nap infants showed semantic priming effect (N400) (usually found later in development and associated with formation of genuine words). This shift was associated with NREM stage 2 length and to the increase in spindle activity. The more time children spent in N2, the higher their N400 was. The N400 effect was correlated with number of spindles in the long-N2 group,

						but not in the short-N2.
Axelsson et al. (2018)	40	2;5	Group assignment was done on the basis of whether children spontaneously napped during the 4h interval between training and test.	Four novel word-object associations (nouns) in competition with 58 familiar words, selected from OZI (Australian adaptation of the CDI).	Learning occurred through referent selection, where eight trials had one novel and two familiar objects, and other eight trials were with familiar words only. Feedback was provided to the children during this phase. At test, the four novel objects were displayed on the screen and children were asked to point to the one they heard. The study comprised three test sessions: immediate, post-nap (after 4h), post-nocturnal sleep (following morning).	In the learning phase through fast mapping both groups performed above chance, with similar accuracy. At immediate test, both groups performed above chance, with same accuracy. After a delay, nappers performed above chance and significantly better than non-nappers. After nocturnal sleep, only nappers performed above chance and significantly better than non-nappers. Nap group's retention scores remained stable, while no-nap group had steady decline in retention scores. Sleep duration positively correlated with retention post-nap.
Munro et al. (2012)	49	33 months	Within-subject design (four tests: immediate,	Eight new words + referent	Didactic training: six exposures to each word and its	Children's recognition of the new words and referents

			after 1 minute, 5 minutes and 4 days)		referent, i.e. an unfamiliar toy.	improved over time. However, children’s production accuracy (at whole-word, syllable and phoneme levels) decreased significantly from one interval to the next.
Dionne et al. (2012)	1029	6, 18, 30 months (initial measures); 18, 30, 60 months (final measures).	Longitudinal, survey-based study	n/a	Language skills measured via CDI & Peabody; Sleep consolidation measured via by the Quebec Newborn Twin Study	From 6 to 30 months, children gradually slept longer and more consecutively at night than during the day (i.e. gradually smaller day/ night sleep ratio). Children with language delays at 60 months had had more immature sleep consolidation at 6 and 18 months.
Horváth et al. (2016)	246	Initial assessment: between 7.73 and 37.83 months	Longitudinal	Sleep and Naps Oxford Research Inventory (SNORI); Oxford Communicative Development Inventory (OCDI)	After an initial assessment of both sleep behaviour and vocabulary, OCDI data were collected again between 2 and 8 times more.	At initial assessment, sleep efficiency (i.e. total time spent asleep/sleep duration) was positively correlated with both receptive and expressive vocabulary, which also correlated with number of naps. The number of daytime naps and sleep efficiency were significant

						predictors of receptive vocabulary (*p* = .006; *p* = .001 respectively), but only marginally significant predictors of expressive vocabulary (*p* = .062; *p* = .068 respectively). The length of nocturnal sleep was negatively correlated with expressive vocabulary development (*p* = .045).