Learning New Words via Reading:

The Influence of Context and Diversity



Matthew HC Mak

Wadham College

University of Oxford

Supervised by
Prof. Kate Nation and Dr Yaling Hsiao

Submitted: Hilary Term 2021

This thesis is submitted to the Department of Experimental Psychology, University of Oxford, in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Declaration

I declare that this thesis is entirely my own work, and except where otherwise stated, describes my own research. In line with open science practice, all the materials, data, and scripts associated with this thesis have been made available on the Open Science Frameworks.

Materials/Data/Scripts from Chapters 2 and 3 can be retrieved from:
https://osf.io/sjwgr/

Materials/Data/Scripts from Chapters 4 to 6 can be retrieved from:
https://osf.io/a57bn/?view_only=160f92cf350c472e8a968bfa0106a6ae

Matthew HC Mak
Wadham College

Table of Contents

Acknowledgements……i

Publications arising from this thesis……v

Abstract……vi

Chapter 1

The effect of various contextual factors on word learning: a review ……1

Chapter 2

What are contextual and semantic diversity?……23

Experiment 1
The influence of semantic diversity on incidental word learning……33

Simulation 1
Simulating spreading activation in lexical networks that differ in semantic diversity (1)……..46

Chapter 3

Simulation 2
Simulating spreading activation in lexical networks that differ in semantic diversity (2)……..53

Experiment 2
Anchoring a word before introducing diversity……59

Chapter 4

Simulation 3
Using word vectors to simulate the learning of meanings across Experiments 1 and 2…….70

Chapter 5

Experiment 3
How does semantic diversity influence online processing? An eye-tracking study……79

Chapter 6

Experiment 4
Can explicit definitions serve to anchor?……97

Chapter 7

General Discussion……119

References……a-v

Appendices……A-G

Acknowledgements

This is an unusually lengthy acknowledgement. Having written this 48,000-word thesis, I feel entitled to write a 2,000-word acknowledgement. There are too many people I need to thank, too many things that I am grateful for, and I would really like to take this opportunity to say thank you to each and every single one of them. They have been instrumental to both my personal and intellectual growth.

First and foremost, my deepest gratitude goes to my mentor, Professor Kate Nation, who has been the best supervisor any graduate students can ask for. If PhD supervision were an Olympic sport, Kate would win the gold medal hands down. The past 3.5 years with Kate have been the most intellectually rewarding period of my life so far. I will always remember feeling awed and inspired by her wisdom and her brilliant mind. What I learned from Kate was not factual knowledge, but the way she thinks, the way she dissects a problem, and the way she handles different matters, from administrative stuff to research. The most impressive of all is perhaps her writing; it's always crisp and always refreshing, often to a point that gives me goose bumps. I have learned *so much* from her magical editing, which has pushed my writing skills to new heights and opened my eyes to my weaknesses as a writer. I will consider myself successful if my writing will ever be as good as Kate's. Another thing that I particularly appreciate about Kate is that she could tear down my mediocre work without making me feel bad about myself. My EPS talks in January 2020/2021 would have been a huge disgrace had Kate not torn down the original version. She did it so gently and delicately that I almost didn't realize she was basically hinting that the draft was awful. This saved me from the dreadful feeling of self-hate and being a complete failure. After tearing down the original version, she offered top-notch advice on how to perfect the presentation, which ended up being a shining moment in my PhD journey. I wouldn't have achieved that or other successes during my PhD if it weren't for Kate's invaluable inputs. Finally, what makes Kate so precious is that she cares not just about your work but also you as a person. Although she often has a hundred balls to juggle at one time, she will always find time to meet with you, and she will remember your achievements as well as all the tiny things that you wouldn't expect her to remember. Having Kate as my supervisor has made my PhD both rewarding and hugely enjoyable. I am forever grateful to her for being the amazing person she is, for taking me under her wings, for all the big and small opportunities along the way, for all the dinners at Branca and SJC, for inviting us to her lovely house to watch the World Cup and to have our fashionably late Christmas Party in February, and of course, for

introducing me to Jonathon. Being Kate's student has been a tremendous honor and privilege.

Second, I would like to take this opportunity to express my heartfelt gratitude to the Drs. Richard Charles and Esther Yewpick Lee Charitable Foundation for their generous financial support throughout the last 3.5 years, without which this PhD would not have been possible in the first place. Not only did the R C Lee Centenary Scholarship cover my tuition fees and my living expenses, but it also supported my undertaking of an MRI course at the JR hospital, my conference trip to Toronto and Montreal, and my vacations in California and Maine. The amount of financial support is unsurpassed, which has enabled me to be a better scientist and to expand my horizon. I am eternally grateful for the Foundation, and I hope I will be successful enough in the future to establish a similar scholarship to pay it forward. I would like to thank the selection committee, particularly Dr Rudgard, Judge Bennett, and Mr. Dockerill, for believing in me and for their friendship and guidance. I would also like to thank Mrs. Lee for being a superb scholarship administrator and for being an amazing friend and listener. Cheers to our shared love for Agatha Christie!

Then, I would like to thank everyone in ReadOxford, past and present. Thank you Yaling for being my second supervisor. Learning with and from you has been a hugely rewarding experience. Thank you for all the buffet lunches at SJC and for being a great friend. I am sure we will have a lot of collaborations in the future, provided that we both stay in academia! Cheers to boba and sweet potato balls! Thank you Nicky for the daily SUAW since March 2020 which helped keep me sane through lockdown! I will always remember our amazing trip to Toronto: the AirBnBs, Bang Bang, and our hike near the "ocean". I hope we will be able to re-create something similar in California with Cam in 2022! Thanks Ellen for all your swimming stories and adventures! Thanks also to Megan, Ascen, and Nick.

Thanks must also go to members of the wider Psychology community here in Oxford, particularly OSCCI. Thank you Dorothy for being Dorothy. Even though I have never been your student, I have learned so much from you about open and reproducible science. Your sagacity, knowledge, and eloquence are stunning. Thank you Adam, Nuala, Paul, and Zoé (in alphabetical order) for being such amazing friends and for sharing my obsession for Pippin doughnuts. You guys made the office a fun and lively place to work at. I really enjoyed our countless conversations over lunch and on Zoom. It's a shame that we can't keep up our tradition of watching the GBBO finale together this year! Anyway, more specifically, I want to thank Adam for teaching me eye-tracking, which made my life 100x easier; Nuala for

being the best person to talk to and her infectious laughs; Paul for all the statistics/RMarkdown advice; Zoé for being an all-round amazing friend and her heavenly cookies. You guys rock!! Thanks also go to Julia for her stories with her cats, for being the nicest person ever, and for inviting me to sit on an interview panel, which turned out to be a life-changing experience. I would also like to thank Professor Kim Plunkett, whose feedback for my transfer and confirmation proved to be extremely useful. Thank you to EP for being home to all the brilliant minds and for all the seminars and teaching opportunities. I would also like to thank (some of) my Cognition and Developmental tutees, the exchange students from the US in particular (Thank you Dr. Dowker for recommending me to Wycliffe!). Teaching has never been so fulfilling. Thank you to the 100+ EP undergraduates who took part in the studies reported in this thesis. Thank you also to Wadham College for being a silently brilliant college and for all the invigilation opportunities.

Next, I would like to thank Sam Davenport for being a great friend and a top-notch squash buddy. You helped keep me sane. Thank you for all the movie nights (Avengers!), your thoughtful birthday gift, Zoom chat, and Catalan games. Also cheers to our love for all-you-can-eat! William Prescott also deserves a big fat thank you. You have been an amazing friend; I just love love love talking to you. So glad that we are always on the same wavelength. Thank you also for all the long walks during the summer of 2020. They were highlights of my times in Oxford. Next on the list are all the Hong Kong peeps. Thank you for all the hotpot, karaoke, and game nights. Finally, thank you FM for our very short story and for making me feel I mattered. I will never forget Montreal and Iceland.

Then, my friends and family in Hong Kong. My sincere gratitude goes to the Tams (TT, Suk Suk, Zoe), who have been instrumental to my growth as a person. I would not have achieved all the things in my life if they didn't nurture me the way they did. Thank you to my mentors at HKU, Professor Stephen Matthews and Dr Kathy Shum. Thank you Jacky for being my best friend, my rock, and for constantly tolerating me. Thank you Rachel (and your family) for being my true friend for 14+ years. Thank you to all my mum's friends who showered me with love during my childhood. Thank you the Bobo family for all the heavenly dinners throughout the years. I am also indebted to my parents, who have provided me with the best possible environment to grow up in. All these would not have been possible without my parents. I know I don't say it often, but I really do appreciate and love you both.

I would like to thank all the freedom-loving Hong Kong people, who showed the most impressive fortitude and courage to protect the place we call home. Your gallantry, perseverance, virtues, compassion, and selflessness are inspiring and have made me so

proud of being a Hongkonger. To all the people who have been standing up for what is right despite the increasing threats: I salute you. Thank you Hong Kong for being the best place in the world to grow up in. Although the Hong Kong as we knew it is gone, I still have hope for this place, because I believe in Hong Kong people. I have no doubt that it will soon restore its dazzling beauty.

If I could go back in time, I would also like to say thank you to my teenage self, for studying English hard, for the vocabulary books you kept, for not losing hope, and for being such a strong, undefeatable young man. I foresee that I will revisit this thesis in the future, so to my 30/40/50/60/70/80/90+ year-old self: I hope you will hold onto science, your beliefs, and the truths. I hope and I pray that you will have the courage to stand up for what is right, the courage to believe in the goods in humans, the courage to stand on the right side of history, the courage to resist temptations, the courage not to sell your soul, but most important of all, the courage to listen to and stand with young people. I would be very disappointed if you become one of those snobbish old men who scorns the vision of young people or one of those horrible human beings who knows no shame and holds double standards. When you look back in time, I hope you will be glad that your 27-year-old self finished this thesis. It is an accomplishment that you should be very proud of. No matter where you are, no matter what job you have landed, I hope you are happy and healthy, are living the life you want to live, and have found what you have wanted to find all along.

March 2021

Publications arising from this thesis

Some ideas in Chapters 2 and 3 appeared as a commentary in *Cognitive Neuroscience*:

Mak, M. H. C. (2019). Why and how the co-occurring familiar object matters in Fast Mapping (FM)? Insights from computational models. *Cognitive Neuroscience*, *10*(4), 229-231. https://doi.org/10.1080/17588928.2019.1593121

Chapters 2 and 3 appeared as a regular article in *Journal of Memory and Language:*

Mak, M. H. C., Hsiao, Y., & Nation, K. (2021). Anchoring and contextual variation in the early stages of incidental word learning during reading. *Journal of Memory and Language*, *118*(June 2020), 104203. https://doi.org/10.1016/j.jml.2020.104203

Abstract

Learning New Words Via Reading: The Influence of Context and Diversity

Matthew HC Mak

From mid-childhood onwards, the majority of new words are incidentally acquired via independent reading. Numerous studies have shown that the linguistic contexts in which a new word appears influence the ease with which the word is processed and acquired. Some contextual characteristics (e.g., high topic familiarity) are more beneficial to learning than others. The focus of this thesis is a relatively new contextual characteristic, known as semantic diversity, defined as the degree to which a word's surrounding contexts overlap in terms of semantic content. The overarching aim of this thesis is to investigate how semantic diversity influences the learning of novel words in young adults. This thesis comprises a total of four word-learning experiments and three computational simulations. They all contrasted two learning conditions: high semantic diversity (multi-themed) and low semantic diversity (single-themed). Both Experiment 1 and Simulation 1 demonstrated that novel words experienced multiple times in one familiar topic (i.e., low semantic diversity) were advantaged in initial learning, a finding attributed to what I refer to as "anchoring", a process of securing novel words onto long-term memory. On the other hand, Simulation 2 and Experiment 2 showed that novel words that had an anchoring opportunity were better placed to take advantage of high semantic diversity, suggesting that anchoring may benefit initial learning. Simulation 3 used word vectors to mathematically simulate how semantic diversity influences the learning of meanings across Experiments 1 and 2. The simulation results showed that high-diversity words tend to have sparser semantic spaces, potentially indicative of weaker integration into pre-existing knowledge. In order to consolidate the results from Experiment 2, Experiment 3 monitored participants' eye movement during reading. The results suggest that novel words appearing in high-diversity contexts are encoded more strongly during reading and tend to have more context-independent representations. Finally, while Experiments 2 and 3 showed that experiencing a novel word multiple times in a repeated topic could serve to anchor, Experiment 4 demonstrated that an explicit definition could also achieve the same effect, enabling the novel words to be better placed to take advantage of high semantic diversity. Together, this research demonstrated that initially restricting a novel word's semantic diversity (i.e., starting small) and capitalizing on learners' prior knowledge can support anchoring, which can facilitate subsequent learning in more diverse contexts.

To my anchors

Chapter 1

The effect of various contextual factors on word learning: a review

1. Introduction

Much research has been dedicated to the question of how to improve reading in childhood and beyond (for a review, see Castles, Rastle, & Nation, 2019). One crucial cornerstone of skilful reading is vocabulary knowledge (e.g., Anderson & Freebody, 1981; Baker, Simmons, & Kame'enui, 1998; Biemiller & Booter, 2006; Hart & Risley, 1995; Muter, Hulme, Snowling, & Stevenson, 2004). The link between vocabulary and reading is almost tautological; after all, one cannot understand a text without knowing the meaning of most words (Nagy, 1988; Simonsen & Singer, 1992). The relationship between vocabulary and reading, however, is not unilateral. Vocabulary growth from mid-childhood onwards is primarily fuelled by reading experience (Nagy & Anderson, 1984): The more a child reads, the larger and richer their vocabulary. The general consensus as to why reading is crucial to vocabulary growth is that it affords people the opportunities to learn new words incidentally from discourse context; that is, new words and their meanings are acquired as people strive to comprehend the texts (Jenkins, Stein, & Wysocki, 1984; Nagy & Anderson, 1984; Nagy & Herman, 1987). Given the reciprocal relationship between vocabulary and reading, literacy researchers have long been interested in how to boost the probability of picking up new vocabulary items during reading. A major strand of research within this theme is to identify what kind of contextual characteristics are beneficial to word learning (e.g., Bolger, Balass, Landen, & Perfetti, 2008; Chilton & Ehri, 2015; Joseph & Nation, 2018; Pagán & Nation, 2019). Before delving deep into this question, we should first consider why it is important to study the role of context in word learning.

Ferdinand de Saussure, one of the most influential linguists of the 20th century, asserted that a word's function and meaning is defined entirely by the word's relationship to other words in the context. Saussure compared this to the state of a chessboard: The value of a chess piece is not determined by where it stands, but by its relative position with other pieces on the board (Saussure, 1916). Words are exactly the same; the meaning of a word is determined largely by its surrounding words in a sentence or passage. Consider *run* in (1) and (2):

(1) Running a marathon is not easy.

(2) Running a company is not easy.

Although the surrounding contexts differ by just one word, the denotation of *run* changes. This illustrates that how words are processed and interpreted is greatly shaped by the context in which they appear. An implication then is that the contexts also influence the ease with which new words are acquired during reading. A large body of evidence suggests that there are different kinds of context and that they are not equal: Some are more beneficial to word learning than others (e.g., Beck, McKeown, & McCaslin, 1983; Hulstijn, 1992). In this review, I summarize and review these contextual characteristics and their influence on word learning. In addition, I attempt to reconcile (seemingly) inconsistent findings in the literature and consider how context can be optimized to promote word learning during independent reading.

1.1. What is word learning and how to measure it?

There are three key dimensions to a written word: phonology, semantics, and orthography. Word learning during independent reading can be understood as the process of establishing robust representation of these three components in long-term memory. Standard laboratory tasks that tap word learning can be broadly divided into two general categories[1]: The first measures how well the word form (phonological or orthographic) is learned; standard tasks include spelling, lexical decision, and orthographic choice. Other tasks aim to index how well the meaning is learned, including meaning matching and meaning generation. Note that the boundary between these two categories is not always obvious as some tasks may measure the learning of both form and meaning (e.g., picture naming). In addition, these tasks vary in terms of their level of difficulty; for example, a spelling task requires significantly more well-specified orthographic knowledge than a lexical decision task (Frith, 1980, 1985). This means that cautions are required when comparing the results of various studies.

If a particular kind of context facilitates word learning, we would expect a boost in performance, relative to a control or a comparison group, on one or some of the aforementioned tasks. However, if a context is facilitative to word learning, it does not necessarily mean that it will boost the learning of both form and meaning; it is plausible that it only facilitates the learning of form but has no, or even an inhibitory effect, on the acquisition of meaning. In this review, and throughout this thesis, I consider the learning of both orthographic form and meaning.

[1] Word knowledge, admittedly, is not only about form and meaning. There are other aspects of word knowledge, such as grammatical functions and syntagmatic association (Webb, 2007). However, it is beyond the scope of this review to examine the learning of all kinds of word knowledge.

2. Contextual characteristics

Drawing across studies that have investigated the interaction between context and word learning, it is possible to identify five major contextual characteristics that influence word learning. Table 1.1 defined these characteristics. Sometimes, their influence is not clear-cut and work in complex ways with other variables. The goal of this chapter is to provide a clear and comprehensive account of how these contextual characteristics modulate the word-learning process. This review draws on a broad literature on word learning from independent reading, mainly from the first language (L1) literature, but informed by studies of second language acquisition (L2) where appropriate, and covers word learning in both children and adults.

Table 1.1

A list of contextual characteristics and their definitions.

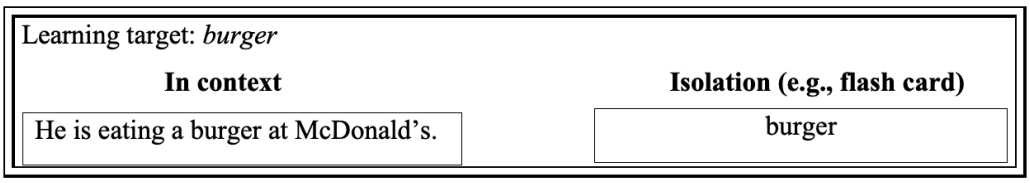
Contextual characteristic	Definition	Binary or Graded?
Contextual presence	Whether the target word was learned in isolation (e.g., flash card) or in context (e.g., story).	Binary
Contextual informativeness (or richness or constraint)	The degree to which the surrounding context reveals the meaning of the target word (e.g., *What is gaichoi?* vs. *Gaichoi is a green leafy vegetable that is very delicious.*).	Graded
Topical familiarity (or prior knowledge)	The degree to which the topic of a text is familiar to the readers. Often operationalized as participants' subjective rating. May also be operationalized as the proportion of unfamiliar words in text.	Graded
Contextual repetition	Whether the context surrounding a target word is identical at each exposure (e.g., identical vs. varied contexts).	Binary
Contextual diversity (or variation) – *To be discussed in Chapter 2*	The degree to which a word's contextual environment overlap. There are different ways to index contextual diversity, including document count (e.g., Adelman, Brown, & Quesada, 2006) and semantic diversity (e.g., Hoffman & Woollams, 2015; Jones, Johns, & Recchia, 2012).	Graded

2.1. Contextual presence

Contextual presence is concerned with whether orthographic learning fares better in context (see left panel of Figure 1.1) or in isolation (see right panel of Figure 1.1). Before addressing this question, it is important to first consider how orthographic learning occurs.

Figure 1.1

An illustration of contextual presence vs contextual absence.



According to the self-teaching hypothesis (Share, 1995) and ample empirical evidence (for a review, see Nation & Castles, 2017), sounding out a word (i.e., decoding) is vital to orthographic learning. That is, if one is able to decode a written word, the person should be able to acquire fairly good—if not perfect—knowledge about its orthography. Studies on the interaction between contextual presence and online decoding consistently showed that decoding accuracy is superior when a target word is encountered in context (vs. isolation; Perfetti & Roth, 1981; Stanovich, 1980; Wong & Underwood, 1996). For example, when given an unfamiliar word, say *burger*, a novice reader is more likely to pronounce it correctly when it is embedded in a meaningful sentence such as "*He is eating a burger in McDonald's*" than when it is presented alone on a flash card. Given decoding is a crucial process in orthographic learning, it seems reasonable to infer that contextual presence may benefit the learning of orthographic form more than when the word was encountered in isolation. However, as multiple experiments with children have shown, contextual presence appeared to have a nuanced effect on orthographic learning: Some reported no discernible effect while some showed an inhibitory effect. What is more interesting is that under some circumstances, contextual presence can facilitate learning.

In Nation, Angell, and Castles, (2007), 6-to-8-year-olds read aloud made-up words (e.g., *ferd*) presented either in meaningful stories [in-context] or one by one on a flash card [isolation]. One and seven days after exposure, the children completed an orthographic choice task in which they chose among four words the one they had previously encountered (e.g., *ferd*, *furd*, *ferp*, *furp*). In both sessions, word targets presented in context were identified equally well as those learned in isolation, suggesting that orthographic learning may not be affected by contextual presence. This null finding was later extended by Wang, Castles, Nickels, and Nation (2011, Experiment 1) who used a spelling test to tap orthographic knowledge.

In contrast to Nation et al. (2007) and Wang et al. (2011, Experiment 1), Landi, Perfetti, Bolger, Dunlap, and Foorman (2006) reported that the presence of context may be *detrimental* to orthographic learning in the same age group (i.e., 6-to-8 years old). In their experiments, children were first asked to read aloud unfamiliar words (e.g., *rose*,

occurrence) in stories or in isolation. One week later, their subsequent success in decoding the words—this time all in isolation—was measured. Here, words previously encountered in context were read aloud significantly *poorer* than those encountered in isolation, suggesting that contextual presence may hinder the long-term development of phonological and orthographic representations. In a similar vein, Martin-Chang, Ouellette, and Bond (2017) showed that in 7-year-olds, target words (e.g., *monster*, *hazardous*) were spelt with lower accuracy in a delayed session if they were previously encountered in stories than in isolation. Their study demonstrated that encountering new words in isolation may produce higher-quality orthographic representations, thereby supporting more precise spelling.

To account for the detrimental effect of contextual presence in the learning of novel word forms, Landi et al. (2006) and Martin-Chang et al. (2017) appealed to the same explanation: The presence of linguistic context channels attention to the semantic content of the discourse, providing contextual cues (top-down support) for decoding the target words. This lessens the probability of children actively attempting to decode the news words using spelling-to-sound correspondences. This diminished focus on the word's orthographic and phonemic properties reduces the likelihood that the orthographic form of the new word is committed to memory. This, in turn, explains why learning of orthographic form may be less satisfactory in context than in isolation.

Not all studies with children reported a negative effect of contextual presence. As noted earlier, Nation et al. (2007) and Wang et al. (2011) found no effect of context. Then, how can we make sense of these seemingly incompatible results? This discrepancy is likely a consequence of different types of target words being used. In studies that yielded a neutral result (e.g., Nation et al., 2007; Wang et al., 2011; Experiment 1), the targets were pseudowords (e.g., *ferd*). On the other hand, studies that reported a negative result (e.g., Landi et al., 2006) utilized words (e.g., *monster, rose*) whose orthographic forms are unknown to the children. For pseudowords such as *ferd*, it is safe to assume that participants had never encountered them prior to the experiment. In contrast, words such as *rose* and *monster* might well be familiar with children having heard them in previous spoken discourse and aware of their meanings. Therefore, before the experiments took place, these words (e.g., *monster*) may already have a phonological and a semantic entry in the child's lexicon, even though an orthographic entry is lacking. When *monster* is read in context, its pre-existing phonological and semantic representations support contextual guessing, and this results in a subpar orthographic representation being built, relative to *monster* being experienced in the isolation condition. On the other hand, when *ferd* is read

in context, the absence of a pre-established phonological and semantic representation lessens contextual guessing and instead effort is allocated to decoding the word form. This allows the orthographic form of the target words to be learned equally well as those encountered in isolation.

Based on the evidence discussed thus far, it is tempting to conclude that contextual presence does not support the learning of new orthographic forms. The full picture, however, is not that straightforward.

As demonstrated by Wang et al. (2011, Experiment 2), under some circumstances, the presence of linguistic context may facilitate learning: In this experiment, 7-to-8-year-olds were first familiarized with the phonology and semantics of pseudowords; then, the children entered a decoding phase where they read aloud the visual forms, which were embedded in either stories [in-context] or wordlists [isolation]. Critically, some of these pseudowords were assigned an irregular spelling: For example, the novel word /kleip/ was given an irregular spelling of *cleap* (a regular spelling would be *clape*). This way, the word's phonology and orthography did not match. Ten days after learning these irregular mappings, the children completed an orthographic decision task where they decided whether /kleip/ is spelt *cleap* or *clape*. Here, pseudowords learned in the story condition were recognized significantly better than those in the isolation condition, leading the authors to argue that the effect of contextual presence on form learning may be modulated by letter-sound regularity (see also McKay, Davis, Savage, & Castles, 2008). Note, however, that the contextual advantage was only observed in orthographic decision and not in spelling.

One way to make sense of this contextual advantage is that learning the orthographic form of irregular words requires some degree of word-specific learning (e.g., rote memorization); using the most frequent letter-to-sound correspondences would guarantee mispronunciations. In Wang et al.'s (2011) study, the irregular words were encountered either in context or in isolation. When they were encountered in context, the discourse environment could serve as a contextual cue (i.e., top-down support) to remind the participants of its meaning and pronunciation, acquired prior to print exposure. [2] This, in combination with a decoding attempt, could bring pronunciation in line with the correct form and so drive orthographic learning. In the isolation condition, there is no such "reminder" from top-down sources to support decoding; instead, children may turn to letter-to-sound conversion, which almost guarantees incorrect decoding. Following the

[2] For example, if children learned that /kleip/ is a fishing machine, a sentence like "He caught a fish with a cleap" may help remind readers of its pronunciation more than encountering *cleap* in isolation, thereby increasing the chance of it being accurately decoded.

notion that successful decoding is pivotal to orthographic learning (Share, 1995; 1999), it is therefore reasonable to see that contextual presence may facilitate the learning of irregular word forms more than contextual absence (Wang et al., 2011; Experiment 2).

Finally, one aspect of the "contextual presence x word learning" literature that merits a brief discussion is the role of semantics. Studies like Landi et al. (2006) and Nation et al. (2007), which I have used to illustrate the effect of contextual presence on orthographic learning, can be interpreted as studies that examined how the provision of semantics influences the learning of word forms. Take Nation et al. (2007) as an example; children were asked to read aloud pseudowords such as *ferd* in a story like this one:

> *Every day after school, Anne took her dog Spot to the park to play. Spot had all sorts of exciting toys. He had a bone which he liked to bury, a rope which he loved to play tug with, and a small rubber* ***ferd*** *which Anne threw for him to catch.* […]

Target words in this story [context] condition had a semantic referent while those in the wordlist [isolation] condition did not. In other words, studies like Nation et al. (2007) were also epiphenomenally investigating whether the provision of semantics influences the learning of orthographic form. Results from these studies suggest that the presence of contexts, and hence the provision of semantics, does not support orthographic learning, at least when the words to be learned are regular words.

When semantics is operationalized differently, however, the picture becomes different. In studies like Nation et al. (2007), the meanings of the target words in the context condition had to be inferred from the story. Other studies have operationalized meaning using explicit definitions (e.g., Ouellette & Fraser, 2009; Ouellette, 2010). In these studies, participants were exposed to target words with a definition or in isolation, as in the example below (Taken from Ouellette & Fraser, 2009):

With semantics

Cale

This is a type of animal that lives in a garden in New York City. It has short curly orange hair and wears a thick silver collar and carries a small bag everywhere it goes to keep its most important possessions close. […]

Without semantics

Cale

These studies have tended to find that the provision of definitional knowledge (vs. its absence) benefits orthographic learning (e.g., Ouellette, 2010; Ouellette & Fraser, 2009; Suarez-Coalla & Cuetos, 2017). For example, in Ouellette and Fraser (2009), target words (e.g., *cale*) trained with an explicit definition were recognized with greater accuracy in an orthographic choice task where 9-year-olds had to indicate among four options (e.g., *cale*, *cail*, *cade*, *cael*) which was the one they had learned during training. The same benefit has also been reported in adults: In Taylor, Plunkett, and Nation (2011; Experiment 2),

undergraduates learned an artificial orthography; before seeing the orthographic form for the first time, half were repeated as isolated phonological forms while the other half were additionally given an explicit definition (e.g., *an assistant to a magician or scholar*). During training, words with a definition were easier to read, and after training, they were recognized with greater accuracy in an old/new judgement task. These suggest that the provision of semantics—when operationalized as definitional knowledge—can support the learning of orthographic forms. One explanation to these findings is that meaning helps hold the phonology of a word together and highlight its specificity (Patterson & Hodges, 1992; see also Taylor, Duff, Woollams, Monaghan, Ricketts, 2015 for further discussion). However, it remains an empirical question as to why meanings derived from contexts does not have the same effect on orthographic learning as definitional knowledge. One possibility is that the former may not be well-specified enough to support learning.

To summarize, this section reviewed the effect of contextual presence vs. contextual absence in the learning of orthographic forms during reading. Existing evidence suggests contextual presence facilitates online decoding, but not necessarily the orthographic knowledge of regular word forms in children (e.g., Landi et al., 2006; Nation et al., 2007), because contextual presence may divert reader's attention away from the word's phonemic and orthographic properties. However, there is some evidence suggesting that the learning of irregular word forms (e.g., *pint*, *yacht*) may be better learned when they are embedded in meaningful contexts than when presented alone (e.g., Wang et al., 2011; Experiment 2). Finally, the availability of meanings also influences orthographic learning, but the direction of effect seems to depend on how meanings were operationalized (i.e., derived from contexts vs. explicit definitions).

2.2. Contextual informativeness/constraint

Contextual informativeness (or contextual richness) refers to the amount of contextual information that gives clues about the meaning of the target words (Haastrup, 1991). Germane to this is *contextual constraint*, which is described as the probability to which the sentential context yields a specific word as the completion of the context (Schwanenflugel & LaCount, 1988). Informativeness and constraint do differ, especially in how they are measured: While informativeness is generally operationalized as subjective rating (e.g., Webb, 2008), constraint is quantified using normative data (e.g., Lowell & Morris, 2017). For the purposes of this review, I group informativeness and constraint together. Below are two

examples illustrating high vs. low informativeness/constraint (van den Broek, Takashima, Segers, & Ludo, 2018).

(1) We do not have any *mkate* left.

(2) I'll go to the bakery; we do not have any *mkate* left.

In sentence (1), the word *mkate* can refer to anything from carrot to flour, from paper to battery; this is because its surrounding context does not constrain its meaning, or in other words, the context is not very informative of what *mkate* denotes. In contrast, the same word in sentence (2) specifically denotes products that can be bought from a bakery, such as bread and croissant. Sentence (2), therefore, can be said to be fairly constraining and informative about *mkate*'s meaning. An important point to note is that contextual informativeness is a graded variable. A context may contain a clear definition of what the target word denotes, while on the other hand, a context may reveal absolutely nothing about the target's meanings. Consider the word *chow mein* in the following sentences.

(3) He eats *chow mein* all the times. *Chow mein* is a Cantonese-style dish of fried noodles with shredded meat or seafood and vegetables.

(4) What does *chow mein* mean?

For (3), the level of contextual informativeness is at maximum: It makes explicit the meaning of *chow mein*. For (4), the level of informativeness is at minimum: The target word can be replaced by essentially any words. Experiments that examined the effect of contextual informativeness/constraint rarely studied the variable at these extremes, as no meaning can be derived from the absolute minimum level; most often, they compared the effect of high and moderate informativeness using sentences like (5) and (6) below (Taken from Webb, 2008).

(5) He could read about things, and talk to his visitors, but he could not go out of the *ancon* by himself. [Lower informativeness]

(6) He was not ill, and of course the beds in the *ancon* are for ill people. [Higher informativeness]

Turning to the influence of contextual informativeness on word learning, my focus will be with learning meaning rather than form. This follows from the observation that contextual informativeness has little impact on orthographic learning (e.g., Hulstijn, 1992; Experiment 5; Lowell & Morris, 2017; Webb, 2008). Its influence has been examined in both the L1 and L2 acquisition literatures and I consider each in turn.

In the L1 literature, there is no clear evidence that the success rate of incidental word learning is affected by whether the context is highly or moderately informative. For example, using a surprise vocabulary test, Diakidoy (1998) reported that high vs. moderate informativeness had no effect on how well the meanings of new words were acquired in a group of 73 sixth-graders. Similarly, Nagy et al. (1985) administered a surprise meaning-matching test (in the form of a multiple-choice test) two weeks after 352 children read novel words in either highly or moderately informative contexts. The meanings of the new words were retained equally well regardless of how informative their contexts were. This null finding with L1 children has also been reported in skilled adult readers (Lowell & Morris, 2017). Taken together, these studies suggest that L1 speakers in general can learn and retain new meaning from texts fairly well regardless of whether the words were encountered in highly or moderately informative contexts.

Although high vs. moderate informativeness does not engender differences in word learning, as assessed by offline tasks in L1 learners, it does influence online processing. In three eye-tracking studies (Chaffin, Morris, & Seely, 2001; Lowell & Morris, 2017; William & Morris, 2014), college students read novel words embedded in either moderately or highly informative contexts. Participants spent more time integrating novel words into the context in the moderate condition whereas in the highly informative condition, they made longer and more frequent refixations on the surrounding context. Despite these differences in online processing, there was no differences in learning after exposure, either in terms of orthographic forms or meaning. This null finding is in line with previous studies with L1 children (e.g., Diakidoy, 1998; Nagy et al., 1985).

While there is no concrete evidence suggesting that highly informative contexts could promote word learning in L1 learners, the same cannot be said for L2 learners. Webb (2008) compared incidental word learning between two groups of Japanese advanced learners of English who first read pseudowords (e.g., *ancon*, which means hospital) in an informative (English) sentence and then in two more sentences that were either highly or moderately informative about the word's meaning. Immediately afterwards, participants were better able to recall and recognize the meanings of the new words encountered in the

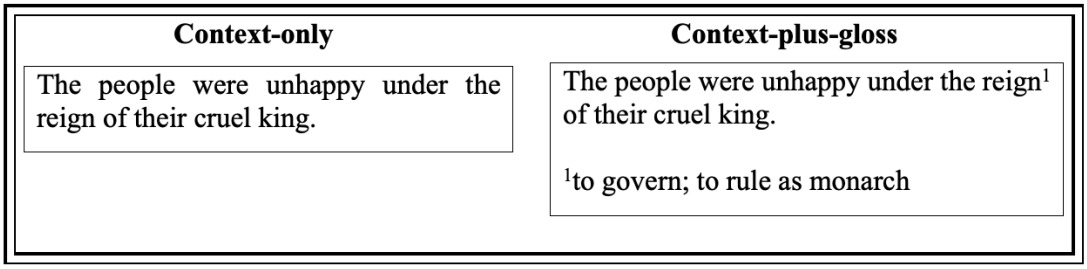
more informative condition. This advantage has recently been replicated in a group of 180 Mandarin-speaking college students, indicating that the advantage is not culture-specific (Teng, 2019). Notably, Teng (2019) administered the post-exposure vocabulary test two weeks after the learning phase, instead of immediately after as in Webb's (2008) experiment. This suggests that high contextual informativeness not just benefits short-term learning but also longer-term retention. Based on these studies, it appears that more informative contexts facilitate the learning of new meanings from reading among L2 adult learners.

In the L2 literature, several studies asked the question of whether *uninformative* contexts could be useful for word-learning. Although these studies were interested in intentional rather than incidental learning, they offer interesting insights into how contextual informativeness can be manipulated in L2 classrooms. For example, in van den Broek et al. (2018), Dutch college students were pre-trained on Swahili nouns alongside Dutch translations (e.g., *mkate* = "bread"). Then, these studied Swahili terms were presented in Dutch sentences, which were either uninformative (e.g., "We do not have any *mkate* left") or highly informative (e.g., "We do not have any *mkate* left. I'll go to the bakery."). Importantly, here, participants were asked to type out the meaning of the target words, in keeping with an intentional learning approach. The researchers predicted that since uninformative context does not reveal much about the target words, learners would be encouraged to actively retrieve the meaning of the target words acquired during pre-training. On the other hand, the highly informative condition would not promote active retrieval as much, since the word's meaning can be easily derived from the context. Given abundant evidence that active retrieval is an efficient way to strengthen memory (for reviews, see Roediger & Butler, 2011; Roediger & Karpicke, 2006), van den Broek and colleagues (2018) predicted that the uninformative condition would endow target words with a higher probability of being retained over time. And this is exactly what their data from three word-learning experiments suggested. They concluded that once L2 learners have some basic understanding of a word's meaning, re-encountering the word in uninformative environments may induce active retrieval. In turn, this helps consolidate learning. Comparable results have also been reported in Dutch children learning French as a second language (Mondria & Wit-de Boer, 1991). Based on the results from these studies, it would be interesting to investigate whether knowledge of newly learned words could be consolidated by the use of contextually impoverished materials (and retrieval prompts) among L1 learners and in more ecologically valid settings.

Finally, relevant to contextual informativeness is a line of studies that investigated whether word learning from independent reading can be enhanced by providing word targets with a gloss or a definition as supplementary materials. This is akin to boosting contextual informativeness by adding an extra sentence to elaborate on the meaning of the target words, but a gloss/definition is often provided in the text margin (See Figure 1.4 for illustration).

Figure 1.4

Context-only vs Context-plus-gloss.



The provision of glosses/definitions has been shown to benefit word learning in L1 learners, although the effect tends to be small and restricted to tasks that tap finer semantic knowledge (Durso & Shore, 1991). For example, Bolger et al. (2008) asked young adults to read texts containing rare words (e.g., *probity*, *wheedle*). Half of these words were given an explicit definition while the other half were not. After reading, participants were able to generate more accurate meanings for words learned alongside a definition. Even though generating a definition is obviously affected by whether or not readers have been exposed to a definition in training, participants only used the words presented in the definition about half the time in the meanings they generated. This suggests that the definitions were not retrieved in verbatim, but "rather interact with memories for sentence contexts" (Bolger et al., 2008, p. 145). However, the same study also showed that in both orthographic choice and sentence completion, providing definitions did not boost performance more than the effect of context alone. Bolger et al. (2008) concluded that while definitions have their values, the scope of their effects appear restricted in L1 adult learners.

Turning to L2 learners, the effect of glosses on word learning seems more favorable, mirroring the findings for contextual informativeness more generally. A large number of studies have demonstrated that L2 learners were better able to learn and retain word targets that were glossed (Hulstijn, Hollander, & Greidanus, 1996, Paribakht & Wesche, 1996; Watanabe, 1997). As to why glosses may be useful, the most apparent explanation is

that they help reduce incorrect inferencing. Inferring word meaning from context does not guarantee that a correct inference is generated (McKeown, 1985); lexical items that are particularly prone to inferencing errors include, for example, homonyms (e.g., *bank*), idioms (e.g., *send to Coventry*), and phrasal verbs (e.g., *look up to*; Laufer, 1997). The provision of glosses, therefore, helps minimize the possibility that a learner arrives at an inappropriate meaning. However, while there is good theoretical and empirical evidence that glosses can benefit the learning of meanings, a few studies reported null results (e.g., Fischer, 1994) and that the benefit could be short-lived (e.g., Jacobs, Dufon, & Hong, 1994). Therefore, glosses should not be seen as a silver bullet to word learning.

To end this sub-section on glosses, I consider what kind of glosses is more beneficial to learning. There are numerous types of glosses: glosses written in L1 and L2, glosses in the form of multiple-choice questions, glosses showing the synonyms of the target words, pictorial glosses, glosses with an audio clip, and so on. A wealth of research in the L2 literature has been dedicated to identifying what kind of glosses is more appropriate for word learning from reading, and below are three preliminary conclusions that can be drawn from the existing literature.

First, there is no clear answer as to whether L1 or L2 glosses are more effective to L2 word learning. Some reported an advantage in L1 glosses (Laufer & Shmueli; 1997), some reported an advantage for the other (Miyasako, 2002); some found no differences (e.g., Jacobs et al., 1994; Yoshii, 2006). One possibility causing such conflicting results is that the level of L2 proficiency may modulate the effect: Miyasako (2002) observed that high-proficiency learners tend to benefit more from L2 glosses while low-proficiency learners are more likely to benefit from L1 glosses.

Second, some evidence indicates that glosses that prompt readers to contemplate on the meaning of the target words (e.g., what is the meaning of *reign*? I. to eat; II. to hide; III. to rule; IV. to announce) facilitate word learning more than those that give away the meaning (e.g., *reign* = to rule) (Nagata, 1999). Lin and Huang (2008) replicated and extended this finding by showing that this benefit is generalizable across L2 proficiency levels. One explanation to this is that glosses that prompt contemplation promote deeper learning, thereby strengthening the memory traces of the target words.

Third, advances in digital media have expanded the possibilities of gloss formats. Glosses are now available in multimedia forms such as pictures, videos, and sounds. Generally, multi-mode glosses (e.g., written definition plus visual information) are more effective in promoting meaning acquisition than single-mode glosses (e.g., written definition

only) (Chun & Plass, 1996; Kost, Foss, & Lenzini, 1999; Plass et al., 1998; Yeh & Wang, 2003). The advantage conferred by multi-mode glosses is usually elucidated within the Dual Coding framework by Paivio (1971, 1991). This proposes that a memory trace is enhanced when it is encoded by multiple perceptual systems (e.g., verbally and visually), compared to when it is encoded by just one system. Although there is much empirical support for multi-mode glosses being more advantageous than single-mode glosses in the learning of meaning, a recent study found that multi-mode glosses (i.e., definition plus picture) led to poorer orthographic learning than definition-only glosses in three separate groups of L2 learners (Boers, Warren, He, & Deconinck; 2017). This is likely because the provision of pictures diverted attention away from the orthographic and phonemic properties of the words, meaning that less decoding took place. This explanation is in essence identical to the one used to explain why orthographic learning might be hampered when new words appear in context than in isolation (e.g., Landi et al., 2006; Martin-Chang et al., 2017).

To wrap up, this section summarized the effect of contextual informativeness, and relatedly, glosses, on the learning of meaning. For contextual informativeness, there is no evidence showing that high and moderate levels of informativeness influence learning outcome in L1 learners, although eye-tracking studies demonstrated a clear difference in how they influence online processing (e.g., Lowell & Morris, 2017). In contrast to the L1 literature, L2 learners seem to benefit more from a higher degree of contextual informativeness, which has been shown to aid both short-term and long-term learning (e.g., Teng, 2019). The precise reason as to why the effect of informativeness differs between speaker groups is an empirical question, but it seems reasonable to assume that it is related to differences in prior vocabulary knowledge. Additionally, some evidence suggests that once equipped with some basic understanding of a new word, L2 learners may benefit from re-encountering the word in uninformative (vs. highly informative) contexts, which may encourage active retrieval (van den Broek et al., 2018). Finally, the provision of glosses is generally favorable to the learning of meaning across L1 and L2 learners, although the effect may be restricted to certain semantic tasks (e.g., Bolger et al., 2008) and may be modulated by the type of glosses used (e.g., Boers et al., 2017).

2.3. Topic familiarity / Prior background knowledge

Topic familiarity—whether a reader is familiar with the topic of a text—depends entirely on their prior knowledge. A topic that is familiar to one person may not be familiar to another.

Studies that examined the role of topic familiarity in word learning can be divided into two categories: (1) whether the learning materials were written in L1 or L2 (L1: Diakidoy, 1998; L2: Kaivanpanah & Rahimi, 2017; Nassaji, 2002; Paribakht & Wesche, 1999), and (2) whether the learners engaged in independent reading or passive listening (independent reading: Diakidoy, 1998; passive listening: Kaefer, Neuman, & Pinkham, 2015). Despite different rationales and experimental designs, the results of these studies paint a consistent picture: Readers who possess more background knowledge of a text's topic generally show more thorough comprehension of and learning from the text. Below I briefly describe some of these studies; then, I will explain why high topic familiarity is beneficial to word learning using insights from both the psycholinguistic and memory literature.

In Diakidoy (1998), 73 sixth-graders read two expository texts containing target words that were unfamiliar to children of that age (e.g., *nomad*, *veil*). The topics of the two texts were desert traders and silk. Before reading these passages, the children were tested on whether they knew a list of "enabling concepts"; for example, the list for the passage on desert traders included *Africa*, *climate*, *oasis*, and *camel*. Diakidoy (1998) hypothesized that having prior knowledge of these enabling concepts should facilitate the incidental learning of the unfamiliar words in the passage (Note: these enabling concepts did not appear in the passage). After completing the test on enabling concepts and reading the passages, semantic knowledge of the unfamiliar target words (e.g., *nomad*) was tested. A multiple regression showed that performance on the enabling concepts test captured a substantial amount of variance in word learning such that students who knew the enabling concepts were more able to learn the target words than those who did not. This indicates that readers are more able to learn the meanings of new words during reading if the discourse topics are more familiar to them.

While Diakidoy (1998) operationalized topic familiarity as knowledge of enabling concepts, it can be operationalized as the density of unfamiliar words in a passage. Generally, the proportion of unknown words in a passage is inversely related to how familiar a reader is with a topic; that is, the more unknown words in a passage, the less familiar one is with the topic (Diakidoy, 1998). Using this operationalization, Nagy et al. (1985) reported that when the proportion of unknown words in a passage increases, the 352 primary schoolers in their study learned significantly fewer words during reading. This provides

support to the notion that the availability of relevant topic knowledge is positively associated with the likelihood of successful word learning.

Turning to the L2 literature, in a series of studies, Pulido (2003, 2004, 2007) demonstrated that gains in semantic knowledge from independent reading tend to be higher when L2 learners read texts of more familiar topics. For example, in Pulido (2004), 35 adult L2 learners of Spanish read four Spanish passages, two describing more familiar scenarios (grocery shopping & doctor visit) and two describing less familiar scenarios (scientific publication & house buying). Familiarity was operationalized as participants' subjective ratings on a questionnaire. Each of these passages contained eight target pseudowords, and participants' knowledge of these words was quantified in two surprise tests that took place two and 28 days after exposure. These tests tapped both the learning of orthographic form (word recognition) and meaning (translation production and recognition). Topic familiarity did not influence performance on either word recognition or translation production, but results from translation recognition indicated that more word meaning was learned from the more familiar passages in comparison to the less familiar passages. Notably, this facilitative effect was still present 28 days later, suggesting that topic familiarity influences longer-term retention.

Given the consistent results reported in the literature, it seems safe to conclude that word learning during independent reading benefits more from high than low topic familiarity. Below, I consider why this is the case.

During reading, both linguistic and conceptual processing work in tandem to enable comprehension. How well a text is understood depends not only on language proficiency, but also on how much pre-existing knowledge the reader is able to apply to the text (Rumelhart & Ortony, 1977; Scha, Bruce, & Polanyi, 1986; Recht & Leslie, 1988; Simonsen & Singer, 1992; Spilich, Vesonder, Chiesi, & Voss, 1979). It is hypothesized that the efficiency of linguistic processing is largely determined by one's language proficiency while that of conceptual processing is by one's familiarity with the topic (Horiba & Fukaya, 2015). On this view, reading ease and comprehension are at the highest level when the reader's language proficiency and their familiarity with the topic are both high.

Greater topic familiarity promotes word learning via at least two mechanisms. First, since reading ease is enhanced by high familiarity, more attention can be diverted from conceptual processing to processing the novel words (Chiesi, Spilich, & Voss, 1979; Horst, 2013), thereby enhancing word learning. Second, the chance of correctly inferring the meanings of novel words is enhanced by high topic familiarity. When the topic is familiar,

the relevant schema could be activated, which serves as the foundation for appropriate estimates of the meaning of the unknown words (Freebody & Anderson, 1983). Similarly, Kintsch (1998, p. 94) proposed that familiar topics create a "perceptual filter" that helps readers distinguish between relevant and irrelevant information, increasing the chance that correct inferences are made.

I propose another cognitive explanation to the advantage that familiar topics bring to word learning. Steyvers and Tenenbaum (2005) postulated that when a new word is introduced into a person's lexicon, it will try to establish itself in the lexicon by building associations with pre-existing words and knowledge. When a new word is learned in more familiar topics, the word will have the opportunity to form association with the familiar topic that has pre-existing representations within the reader's long-term memory. Such a linkage with familiar topics may allow the retention of the novel word (both the word form and meaning) to be more reliant on long-term memory (Mak, 2019). In turn, this should help secure and bolster the word's memory representation. In contrast, when the topic is unfamiliar, the novel word is less likely to form associations with representations in long-term memory, not least because such a representation may not exist. Therefore, the resultant representations of the novel word may be weaker or less accurate when it was experienced in unfamiliar topics, relative to when it was encountered in more familiar topics. This proposal will be revisited in Experiment 1.

Finally, relevant to topic familiarity is genre, which is generally divided into two categories: narrative and exposition (Shefelbine, 1990). As far as I am aware, there are only two published studies that explicitly examined whether genre affects word learning during independent reading (i.e., Nagy et al., 1985; Shefelbine, 1990).[3] Nagy et al. (1985) randomly assigned over 300 primary school students from different grades to read either a spy narrative or an exposition on river system. Novel words in both types of text were learned equally well. Shefelbine (1990) also examined the effect of genre on incidental word learning. In his study, 32 sixth-grade students read both a narrative on a girl trying to escape and an exposition on animal. After reading the passages, the children completed a multiple-choice quiz that tested their semantic knowledge of the target words. Although there was no main effect of genre on how well the target words were learned, there was a significant interaction between genre and prior vocabulary knowledge such that children with limited vocabulary knowledge learned fewer target words from the expository texts. Shefelbine

[3] A few studies have examined the effect of genre on word learning during *shared* reading. These studies are out of the scope of the current review, but interested readers can refer to Anderson, Anderson, Lynch, and Shapiro (2004) and Bergman, Deitcher, Johnson, and Aram (2019).

(1990) argued that this may be related to the fact that these children had difficulty understanding common explanatory words—words used to define more difficult words. In turn, this rendered the learning of the target words more challenging in expository texts. It was therefore suggested that pre-teaching these common words may alleviate the word-learning difficulty experienced by students with limited vocabulary knowledge.

As far as I am aware, the two studies discussed above are the only published studies that examined the effect of genre during independent reading. Due to limited evidence, it is not possible to draw any firm conclusion. However, I propose that any potential learning difference between narrative and expository texts may boil down to topic familiarity. Exposition often introduces readers to new concepts, so it is likely that an exposition (vs. narration) requires more conceptual processing, making word learning more challenging.

To wrap up, this section reviewed the effect of topic familiarity on word learning during independent reading. It is clear from the existing evidence base that high topic familiarity facilitates word learning, regardless of whether it is L1 or L2 learning that is at stake. Explanations to this advantage include: (1) the chance of correct inferencing is enhanced when readers are more familiar with the topics, (2) high familiarity increases the ease of conceptual processing, enabling more attention to be allocated to word learning, and (3) the learning of new words may become more reliant on long-term memory when topic familiarity is high. Finally, the effect of genre (exposition vs. narration) on word learning may boil down to differences in how familiar the readers are with the topics.

2.4. Contextual repetition

Contextual repetition refers to whether a word is experienced in identical context. Consider the word *wheedle* in pairs (1)-(2) and (3)-(4), taken from Bolger et al., 2008:

(1) Many wives can wheedle their husbands into things they don't want to do.
(2) Many wives can wheedle their husbands into things they don't want to do.

(3) Many wives can wheedle their husbands into things they don't want to do.
(4) If you carefully wheedle them, they will think they are getting their way even while working for you.

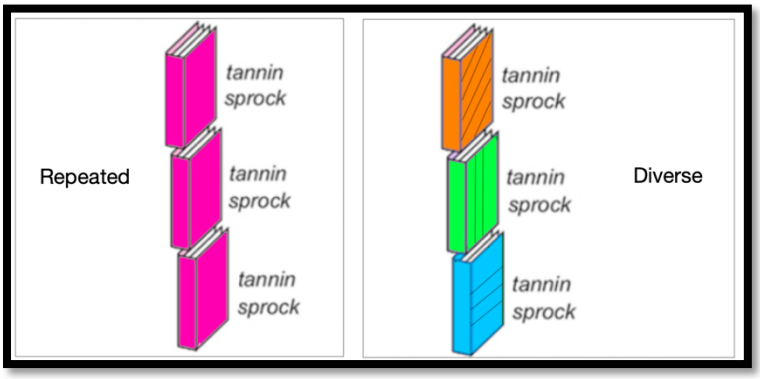
In the first pair, the context for *wheedle* was repeated, while in the second pair, the surrounding context changed. A few studies (e.g., Bolger et al., 2008) have asked whether

unfamiliar words like *wheedle* would be better learned when the context is repeated or varied. The answer to this is not straightforward.

One factor that influences the effect of contextual repetition might be age. In a series of studies, Jessica Horst and colleagues examined the effect of contextual repetition on word learning from context in toddlers. It is important to note that these developmental studies examined word learning via shared reading (i.e., children listened to stories read aloud by caregivers/researchers), so these studies should be considered peripheral to the current review. Despite this, they offer some interesting insights into the effect of contextual repetition in early word learning. Horst, Parsons, and Bryan (2011) assigned 3-year-olds to either the repetition group, where they heard one story three times, or the diverse group, where they heard three different stories (see Figure 1.5 for visualization). Each story contained two target pseudowords (e.g., *tannin* & *sprock*) that were mapped onto made-up objects. Importantly, all children, regardless of group assignment, heard each novel word the same number of times (N = 3).

Figure 1.5

Repetition vs. Diverse conditions in Horst et al. (2011).



Immediately after listening to the stories, children were tested on how well they had learned the object-word referents using a four-alternative forced choice test; testing was repeated one week later. There was no significant difference in recognition accuracy in the immediate test. However, one week later, children in the repetition group outperformed those in the diverse group by a large margin: While the diverse group performed at chance level (25%), the repetition group achieved close to 70% accuracy [$t(14) = 5.25$, $p < .001$, $d = 2.62$]. This finding was replicated and extended by Axelsson, Williams, and Horst (2016) who added a sleep manipulation to the design.

The benefit associated with contextual repetition reported by Horst and colleagues has been extended to independent reading in children. Tamura (2015) reported that 9-to-10-year-olds who re-read the same story were more able to define the word targets compared to their peers who read the same target words equally often but in different stories. While repetition conferred a benefit on the learning of meaning, it did not seem to boost the learning of orthographic form, as children in both the repetition and diversity groups performed equally well on a range of orthographic tasks (i.e., orthographic decision and spelling). Together, these studies provided evidence that contextual repetition may aid the learning and retention of novel word meanings in toddlers and children. A possible explanation to these findings is that contextual repetition helps children to allocate fewer cognitive resources to conceptual processing (e.g., the storyline), enabling more resources to be allocated to the novel aspects of the stories, that is, the new words (Horst, 2013).

The benefits associated with contextual repetition, however, were not seen in a word-learning experiment with young adults (Bolger et al., 2008). In this study, undergraduates first read low-frequency English words (e.g., *assiduous*, *fribble*, *wheedle*) presented either four times in the same sentence (repetition) or in four different sentences (diversity). In the post-tests, target words encountered in the diverse condition were more likely to be recognized in an orthographic choice task (Experiment 1) and be defined more accurately in a meaning generation task (Experiment 1 & 2), suggesting that contextual repetition (vs. diversity) reduces the learning of both orthographic form and meaning in young adults, relative to contextual diversity.

In an eye-tracking study, Pagán and Nation (2019) examined the effect of contextual repetition on word learning. As in Bolger et al. (2008), low-frequency words (e.g., *probity*) were presented in carrier sentences that were either identical on each reading (repetition) or varied in content each time (diversity). In the post-test, participants read the target words in "neutral" sentences, low in contextual informativeness. Pagán and Nation (2019) reported that although target words in the diverse condition were fixated at longer during the exposure phase (suggesting more effortful processing), they were fixated *shorter* in the post-test, suggesting that the adult readers "had less difficulty identifying and integrating the newly learned diverse words in new contexts than words they had experienced equally often but in repeating contexts" (p. 18). Taken together, these studies indicate that contextual repetition, unlike in children, does not benefit word learning during independent reading in adults.

Then, why does the learning of new words fare better in diverse contexts among adults? There are multiple equally plausible explanations. First, some memory (Nelson & Shiffrin, 2013) and reading (Reichle & Perfetti, 2003) theories asserted that word knowledge comprises not only the word itself but also its surrounding context. Always experiencing a word in the same context will result in word knowledge that is strongly associated with the context (i.e., highly context-dependent); if this word appears in a new context, readers may experience difficulty identifying and processing the word as the current context does not match the established memory trace. In contrast, experiencing a word in diverse contexts will result in word knowledge that is more context-independent (or decontextualized). Higher context independence, according to Bolger et al. (2008), may allow a word to be more easily identified and retrieved, increasing its readiness to be integrated into new contexts. Diverse contexts, according to the Lexical Legacy Hypothesis (Nation, 2017), has been proposed as a key mechanism to drive word learning from reading, and more generally, has been offered as an organizing principle of lexical knowledge (Jones et al., 2012). Second, and somewhat relatedly, the diversity advantage may be due to the fact that experiencing a word in an array of contexts will endow the word with more retrieval cues, enabling swift and reliable access to word knowledge (Beck, Perfetti, & McKeown, 1982; see also Carey, 1978). Third, a potential contributing factor to the diversity advantage in adults is that varied contexts offer novelty and encourage attention, and from this, further enhances learning. In the next chapter and for the rest of this thesis, I will focus on contextual diversity and examine how it influences word learning from independent reading.

3. Limitations and conclusion

One key limitation of this review is that all contexts were treated as if they were the "same" across studies; however, how "contexts" were operationalized varied from study to study. For example, some operationalized context as a novel (Godfroid et al., 2018; Saragi et al., 1978), some as paragraphs (Jenkins et al., 1984), as narrative and expository texts (Nagy et al. 1985), as multiple (Lowell & Morris, 2017) or a single sentence (Chilton & Ehri, 2015; Pagán & Nation, 2019; Webb, 2007). Speculatively, the success rate of word learning might be inversely related to the number of words in text: The more words a text contains, the less likely a target word is learned. This speculation is based on the assumption that having more words in text entails more conceptual processing, effectively taking readers' attention away from the novel words. However, this inverse relationship may be modulated by a third factor such as age and topic familiarity. Empirical work is needed to pinpoint how text length

interacts with various contextual characteristics to influence word learning. Furthermore, as noted earlier in this review, there are numerous word-learning tasks, from lexical decision to meaning generation, from orthographic decision to picture naming. Not only do they differ in what they putatively measure (i.e., orthographic vs. semantic knowledge), they also differ in their level of difficulty (e.g., a meaning generation task requires more well-specified semantic knowledge than a meaning matching task). The fact that different word-learning measures were used across studies means that cautions are required when comparing their results. One approach to this issue is to conduct a series of studies that employ the same experimental measures, an approach adopted in this thesis.

In this chapter, I reviewed how various contextual characteristics (e.g., contextual informativeness, topic familiarity) influence the ease with which new words are acquired during independent reading. Most contextual characteristics have a nuanced effect on word learning. For example, (1) contextual presence facilitates online decoding but not necessarily the learning of orthographic form, (2) highly and moderately informative contexts appear to have little effect on word learning among L1 learners, but L2 learners showed more semantic learning from highly informative contexts, and (3) contextual repetition seems to favor word learning in young children but not in adults. All these demonstrate that while context clearly influences the ease of word learning, its effect may be modulated by a third factor (e.g., L1 vs. L2 speakers; age). The only contextual characteristic that has a consistent and reliable effect is topic familiarity such that more familiar topics are generally more favorable to the learning of meanings.

To sum up, research to date has showed that learning words from independent reading can be facilitated with the appropriate contextual elements. Language teachers and curriculum planners should therefore pick reading materials with care and provide students with the necessary support (e.g., pre-teaching enabling concepts, providing glosses). As a concluding remark, I borrow Saussure's (1916) view once again: A word's value is determined by its surrounding context, just like a chess piece on a chessboard. It is hoped that this review has provided readers with a better understanding of what chess layouts are more likely to help learners to win the word-learning game.

Chapter 2

What are contextual and semantic diversity?

1. Introduction

One of the reviewed contextual characteristics in Chapter 1 is contextual repetition (e.g., Bolger et al., 2008; Horst et al., 2011). Its opposite is *contextual diversity* (or contextual variation)—a graded variable that is concerned with the extent of variation between a word's contextual environments. An increasing body of evidence suggests that contextual diversity influences the organization of the mental lexicon (e.g., Adelman et al., 2006) and the ease with which new words are learned from reading (e.g., Frances, Martin, & Duñabeitia, 2020; Rosa, Tapia, & Perea, 2017). This chapter begins with how contextual diversity became a recognized variable in psycholinguistics, followed by a detailed discussion on its influence on lexical processing and learning. The next section (Section 2) describes the first experimental study (Experiment 1) of this thesis.

1.1. Contextual diversity: theoretical background

To trace the history of contextual diversity, it is important to first consider word frequency, which is calculated by counting how many times a word occurs in a language corpus. Frequency is a well-loved variable in psycholinguistics, as it is an excellent predictor in a wide range of tasks, from lexical decision (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Forster & Chambers, 1973) to serial recall (Roodenrys, Hulme, Alban, Ellis, & Brown, 1994; Mak, Hsiao, & Nation, 2020), from word naming (Balota et al., 2004) to semantic decision (Pexman, Heard, Lloyd, & Yap, 2017). These findings have made *the principle of repetition* ubiquitous in cognition (Jones, Dye, & Johns, 2017) and have formed the basis of many models concerned with language and cognitive processing (e.g., threshold activation models; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). To instantiate the principle of repetition, these models hold that each encounter with a word strengthens its memory trace, which, in turn, enhances the ease with which the word is processed.

Towards the end of the 20th century, advances in computing capability gave birth to high-quality language corpora (e.g., British National Corpus) as well as more complex techniques in corpus analysis (e.g., inverse document count; Church & Gale, 1995), enabling researchers to examine language corpora with enhanced efficiency and to move beyond counting frequency of occurrence. As psycholinguists gain more refined insights into how the mental lexicon is structured as reflected in language corpora, questions have been raised

about the nature of frequency in lexical processing (for reviews, see Baayen, Milin, & Ramscar, 2016; Brysbaert & New, 2009). A burgeoning body of evidence has suggested that the frequency effect in the psycholinguistic literature may be better explained in terms of contextual diversity, which is conceptualized to reflect how varied a word's contextual usage is (e.g., Adelman et al., 2006; McDonald & Shillcock, 2001). According to Jones et al. (2017), if frequency is based on the principle of repetition, then contextual diversity is based on *the principle of likely need*, which asserts that a word that has been encountered in more varied contexts is more likely to be needed in future unknown contexts, making it more accessible in semantic memory (Anderson & Milson, 1989; Anderson & Schooler, 1991). The principle of repetition and the principle of likely need make different empirical predictions. To illustrate, consider words A and B in Table 2.1.

Table 2.1

An illustration of contextual diversity vs. frequency.

Context	**Content** (each letter presents one unique word)
1	A B C
2	A A B B
3	A B C A A
4	B
5	B D D D

Word	**Contextual Diversity**	**Word Frequency**
A	3	6
B	5	6
C	2	2
D	1	3

In this example, word A appears in three different contexts (1,2,3) and a total of six times. Word B occurs for the same number of times, but appears in five different contexts (1,2,3,4,5). The principle of repetition predicts that the two words should be equal in strength and availability, so there should be no difference in how efficiently they are processed in psycholinguistic tasks. In contrast, the principle of likely need predicts that word B should be processed more efficiently than word A because the former was experienced in more diverse contexts. Adelman et al. (2006) showed that contextual diversity predicts reaction time better than word frequency in both word naming and lexical decision (see also McDonald & Shillcock, 2001; Perea, Soares, & Comesaña, 2013; Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008). The converging results from these studies have

been taken to indicate the possibility that the mental lexicon is organized based on a word's contextual diversity, with words experienced in more diverse contexts being more accessible. This has important implications for our understanding of lexical processing and learning, as contextual diversity signals a need to depart from a learning mechanism that puts repetition at its heart (Perea et al., 2013).

While the principle of likely need provides a theory as to *why* contextually diverse words are accessed more efficiently, it doesn't explain *how* this processing advantage comes to be. One learning model that attempts to bridge this gap is the Semantic Distinctiveness Model (Jones et al., 2012), which postulates that when a word is experienced in context, both the word and its surrounding contexts will be encoded in a memory vector. When the word is subsequently experienced in another text, it will be compared to the stored memory vector. If the context of the new encounter is distinct from the word's stored contextual history, it will exert pressure on the model to update the word's representation. In contrast, if the context is similar to or is the same as what is stored in the memory vector, less updating occurs. The extent to which a word is updated on each encounter depends therefore on the degree of contextual change (see also Nation, 2017). Over time, those words experienced in more diverse contexts may be represented more strongly and become more context-independent (i.e., less associated with a specific context). A behavioral marker of this, according to the Semantic Distinctiveness Model, is faster lexical access. Evidence from behavioral and computational studies have converged to support the model as a potential mechanism by which contextually diverse words gain greater accessibility (for review, see Jones et al., 2017).

1.2. Contextual diversity: methodological concerns

Although it is now generally accepted that the contexts in which a word has been experienced influences its processing efficiency, questions remains as to how a word's contextual diversity is best operationalized. Measuring a word's frequency is relatively straightforward: Simply count how many times a word occurs in a language corpus. The same, however, cannot be said for a word's contextual diversity. This is mainly because there is no general consensus as to what constitutes a context (Hollis, 2020; Jones et al., 2017). A single sentence? A paragraph? A document? Or the spatial-temporal context? This is similar to the issue noted in Chapter 1: Word-learning studies are inconsistent in how they operationalized contexts, some as a book chapter, some as a paragraph, and some as a single sentence. One of the earliest approaches to contextual diversity is Adelman et al.'s

(2006) metric, known as *document count*. It considers a word's context as the document in which the word occurs. A word's document count, therefore, is simply the total number of unique documents the word occurs in a corpus. This means that a word is more contextually diverse if it appears in more documents. As noted earlier, Adelman et al.'s metric has been shown to predict lexical decision and naming latencies better than frequency does. This metric, despite its outstanding predictive power, has been criticized on three grounds. First and perhaps the most important criticism is that document count is so tightly correlated with word frequency ($r > .95$) that it is hard to argue the two variables are genuinely distinct from each other (Brysbaert & New, 2009).

Second, document length often varies substantially across documents (Hollis, 2020). For instance, a document can be a tweet containing fewer than 30 words, but it can also be a news article containing anything between 100 and 2,000 words. In Adelman et al. (2006), the standard deviation of document length (based on the British National Corpus) is 25,914 word tokens. Such a big variation has led to the question of whether document count is a psychologically valid conceptualization of context (Jones et al., 2012), and an increasing body of evidence suggests that a word's context may be better operationalized as a word window of e.g., four words in front of and behind a target word (Hollis, 2020; McDonald & Shillcock, 2001).

The third criticism of document count as a measure of contextual diversity is that it ignores the *semantic content* of a document (Jones et al., 2012; see also Hollis, 2020). For instance, a word appearing in two separate documents about Donald Trump will have a document count of two, but so will a word appearing in two documents, one of which about Boris Johnson, another of which about vegetarianism. Although the principle of likely need as outlined by Anderson and Milson (1989) is indifferent to this issue, Jones et al. (2012) argued that a word's likely need should at least be partially determined by the *content* of the contexts, such that experiencing a word in more diverse *content* will increase a word's likely need to a greater extent than an equal number of exposure in overlapping or highly similar content. To test this hypothesis, Jones et al. (2012) developed a continuous measure of contextual diversity, referred to as *semantic diversity* (originally called *semantic distinctiveness*). Similar to document count, semantic diversity quantifies variation in a word's contextual experience, but unlike document count, it also takes into account how much semantic overlap there is between documents. Semantic overlap was quantified as the proportion of overlapping content words between a pair of documents, based on the rationale that two documents with more overlapping content words are likely to be more

similar to each other semantically. Then, this measure of semantic overlap was averaged across all the document pairs and was used to adjust a word's document count. This means that if two words have the same document count but one appears in more semantically diverse documents, it will receive a higher diversity value because its document count is adjusted upwards due to greater difference between the contents of the documents. Jones et al. (2012) demonstrated that this diversity measure captured more variance in visual lexical decision and word naming in mega-study datasets (Balota et al., 2007) than did either frequency or document count. This finding has been replicated and extended to both bilingual adults and to auditory word recognition (Johns, Gruenenfelder, Pisoni, & Jones, 2012; Johns, Sheppard, Jones, & Taler, 2016).

Another related metric, also called *semantic diversity*, was developed by Hoffman, Lambon Ralph, and Rogers (2013). While this metric also quantifies how semantically similar a word's contextual environment is, its rationale and computational method differ from those of Jones et al.'s (2012) metric. First, Jones et al. devised their measure as a substitute to document count, but Hoffman et al. sought to develop semantic diversity as an independent measure that can capture polysemous variation in word meaning as a continuous, graded property. The second difference is how semantic diversity was computed: Jones et al. used the proportion of overlapping content words between documents to index semantic overlap whereas Hoffman et al. derived vectors for the contexts in which a word occurs using Latent Semantic Analysis (Landauer & Dumais, 1997). To put it in less technical terms, Jones et al.'s semantic diversity is a word-to-context measure, comparing word and context representations, whereas Hoffman et al.'s semantic diversity is a context-to-context measure, comparing contextual representations with each other (Cevoli, Watkins, & Rastle, 2020).

Given these fundamental differences, it is necessary to ask if the two semantic diversity metrics are capturing the same underlying construct. The first step might be to see how well they correlate with each other. Unfortunately, however, Jones et al.'s (2012) metric is not publicly available. Nevertheless, two lines of evidence suggest that these metrics, despite sharing the same name, measure different (although potentially closely related) underlying constructs. First, Jones et al.'s measure is highly correlated with frequency ($r \approx 0.80$) whereas Hoffman et al.'s is only moderately ($r \approx 0.50$). Second, while Jones et al.'s metric captures all variance predicted by frequency in word recognition, Hoffman et al.'s appears to contribute to word recognition independently of word frequency (e.g., Cevoli et al., 2020; Hsiao, Bird, Norris, Pagán, & Nation, 2020; Hsiao & Nation, 2018).

Note also that a recent scrutiny of Hoffman et al.'s metric showed that it may be better named *textual diversity*, because it appears to capture general information about the document's topics, instead of variation in a word's meanings across contexts (Cevoli et al., 2020). I found Cevoli et al.'s (2020) proposal appealing. However, I decided to retain the use of "semantic diversity" here, because how contextual variation was operationalized in the experiments reported in this thesis fitted with both Hoffman et al.'s and Jones et al.'s metrics.

1.3. How does semantic diversity influence lexical processing?

Regardless of which diversity metric is adopted, semantic diversity values derived from language corpora have been shown to reliably predict visual word recognition such that words higher in semantic diversity tend to be recognized faster and/or more accurately in lexical decision and naming (e.g., Cevoli et al., 2020; Hoffman & Woollams, 2015; Hsiao & Nation, 2018; Jones et al., 2012). This finding aligns with the predictions of the Semantic Distinctiveness Model (Jones et al., 2012), which posits that words high in semantic diversity have more accessible and context-independent representations in the mental lexicon. Notably, the effect of semantic diversity cannot be explained by frequency, document count, or other variables that are known to influence word recognition (e.g., imageability), indicating that the effect of semantic diversity is not simply a frequency or a document count effect in disguise.

On the other hand, while high semantic diversity is associated with more efficient visual word recognition, it appears to be detrimental to psycholinguistic tasks requiring semantic judgement, such as relatedness judgement where participants are asked to decide for example whether *chance* and *opportunity* are related in meaning (e.g., Hoffman & Woollams, 2015; Hsiao et al., 2020; Johns, Jones, & Dye, 2016). Potentially, high-diversity words are disadvantaged in this kind of tasks because their meanings might have been extended to fit more diverse contexts (Bolger et al., 2008), making them more semantically nuanced. A consequence of this might be lower semantic stability, which may, in turn, make semantic judgement more challenging (Johns et al., 2016). Chapman (2019) argued that the effects of semantic diversity might be considered in the same way as the effects of semantic richness (e.g., semantic ambiguity; Rodd, Gaskell, & Marslen-Wilson, 2002), which appear to be beneficial to tasks where specific details about a word are not required (e.g., in word recognition) but detrimental to tasks where a specific semantic detail of a word must be identified (e.g., semantic judgement; Hino, Pexman, & Lupker, 2006; Pexman et al., 2008;

Yap, Tan, Pexman, & Hargreaves, 2011). This proposal, however, does not align with the results from Cevoli et al. (2020), who showed that Hoffman et al.'s diversity metric bears no relationship with semantic ambiguity effects in word recognition. Given this, I consider semantic diversity as a separate, albeit potentially linked, construct to semantic richness in this thesis. Future work is needed to further delineate how semantic richness is related to contextual experience.

1.4. Why is it important to study semantic diversity in word learning?

Although semantic diversity values derived from language corpora appear to consistently capture performance in word recognition and semantic judgement, there are two important gaps in knowledge. First, Hoffman et al.'s semantic diversity correlates moderately (*r*s > ± .40) with a range of other lexical variables, including, but not limited to, frequency, age-of-acquisition, and concreteness. While it is possible to regress these factors out in statistical analyses, the semantic diversity effect in lexical processing could well be attributed to some other factor, correlated with semantic diversity but not controlled in analysis (Pagán, Bird, Hsiao, & Nation, 2020). Second, a person's contextual experience with a word is unique, depending on factors such as education level. A word's semantic diversity value is derived from a corpus containing many thousands of texts (e.g., British National Corpus), which does not necessarily reflect the person's language experience. Relatedly, a word's diversity value differs significantly depending on what type of corpus was used (e.g., spoken vs. written corpus, children vs. adult literature; Hsiao & Nation, 2018; Johns, Dye, & Jones, 2020). This in itself is not a limitation, but if semantic diversity really exerts an influence on lexical processing and learning, its effect should be clearest when a word's diversity value is tailored to a person's language experience. One way to address these gaps in knowledge is via word-learning experiments where new words are encountered in different contexts and the effects of this on future processing is assessed. This methodology allows precise control of lexical statistics (e.g., frequency, age of acquisition) and the nature of the contextual encounters. It also offers the potential to track learning over time, as knowledge accrues with successive encounters.

A small number of word-learning studies have taken this prospective approach to study semantic diversity. Unfortunately however, and in contrast to the consistent evidence emerging from studies using corpora-derived diversity values, their results are far from consistent. This is perhaps not surprising as each induced diversity in different ways, indexed

learning using different tasks, and used different participant groups (undergraduate students vs. school children).

Johns et al. (2016) investigated how semantic diversity influences word learning from independent reading in young adults. The researchers first selected 10 low-frequency target words (e.g., *constellation*, *latent*, *libertine, silo*) and developed two sets of passages for each target: One set comprised five passages on one single topic (low semantic diversity) and the other set comprised five passages spanning five distinct topics (high semantic diversity). The target words were then replaced by pseudowords (e.g., *constellation* → *covella*). In the experiment, five of the pseudowords were encountered in the low-diversity passages, and the other five in the high-diversity passages. Post-exposure, participants completed two tasks: (i) speeded recognition (akin to lexical decision), where the participants judged whether a pseudoword had been encountered in any of the passages, (ii) semantic similarity judgement, where the participants rated how semantically similar a pseudoword is to a familiar word . The researchers found that while newly-learned words experienced in the high-diversity condition were identified faster and more accurately in speeded recognition, they were processed less well in semantic similarity judgement. These observations fitted with findings from computational simulations (Jones et al., 2012) and studies that examined the effect of corpus-derived semantic diversity (e.g., Hoffman & Woollams, 2015; Hsiao & Nation, 2018), which showed that high-diversity known words tend to be more accessible in the lexicon but have more nuanced semantic representations that may hinder performance on judgement tasks that require a specific semantic detail.

The effect of semantic diversity on word learning has also been investigated in children. Using a similar design to Johns et al. (2016), Rosa et al. (2017) asked 8-9 year-olds to read unfamiliar concrete nouns (e.g., *canines*, *russet* in Spanish) embedded in either same-themed or different-themed passages. Afterwards, words experienced in different-themed passages were processed more efficiently in a range of tasks. Like Johns et al. (2016), there was a diversity advantage on tasks tapping the learning of orthographic form (e.g., free recall); unlike Johns et al., however, there was also a diversity advantage on tasks tapping knowledge of word meaning. This seemingly discrepant finding is likely due to the fact that Rosa et al. (2017) only used concrete nouns as the word targets and a semantic task (i.e., picture-word matching) that does not necessarily require a specific semantic detail of a word.

In contrast to these findings, Joseph and Nation (2018) reported a null effect of semantic diversity on word learning. In their study, children aged 10 to 11 read sentences

containing unfamiliar verbs (e.g., *exacerbate*, *amalgamate*) while their eye movement was tracked. Similar to the previous two studies, a target word was experienced either in sentences centered on the same topic (low semantic diversity) or in sentences spanning different topics (high semantic diversity). Following the reading phase, the children completed a range of post-tests tapping their orthographic and semantic knowledge of the target verbs (e.g., spelling and plausibility judgement). Performance on the post-tests was equivalent between the two diversity conditions, which also did not differ in terms of online processing, as indexed by the children's eye movements during the reading phase.

Finally, Chilton and Ehri (2015) reported that the meanings of new words were learned less well in high-diversity contexts, a finding that contrasts with Rosa et al. (2017). In their study with 9-year-olds, six pre-trained unfamiliar verbs (e.g., *restrain*, *anticipate*) appeared in sentences that together constituted one single theme (low semantic diversity) or in sentences that each has its own theme (high semantic diversity). Note, however, training involved the experimenters reading aloud the sentences as children viewed them in print. Unlike the previous studies, decoding might not have taken place. One day after exposure, the participants completed a range of orthographic (e.g., spelling) and semantic tasks (e.g., sentence generation, definition production). Results indicate that while diversity did not influence orthographic learning, it had a significant effect on almost all the semantic tasks such that children who encountered the target verbs in the low-diversity condition outperformed their counterparts who encountered the same verbs in the high-diversity condition.

In summary, as Chapter 1 and this section showed, various contextual characteristics influence the ease with which new words are processed and learned. It is clear from corpus works that contextual variation, as captured by semantic diversity, influences lexical processing (e.g., Jones et al., 2012; Hoffman et al., 2013; Hoffman & Woollams, 2015). However, there is a clear need for data from learning experiments to understand its nature. Yet, the existing evidence base is small, methodologically varied and contradictory in nature. With this as a backdrop, this thesis explored how variation in context, as indexed by semantic diversity, influences the learning of novel words during independent reading in young adults. The research reported in this thesis aims to address these research questions:

(a) How does novel word knowledge grow as a function of contextual variation, as indexed by semantic diversity?

(b) How does semantic diversity influence the learning of orthographic form and meaning?

(c) Do other contextual characteristics such as topic familiarity and definitional knowledge influence the effect of semantic diversity?

1.5. Thesis overview

This thesis comprises four word-learning experiments with young adults, complemented by three computational simulations. All these investigations contrasted two learning conditions: high semantic diversity (multi-themed) and low semantic diversity (single-themed). All experiments had semantic diversity as a within-participant manipulation and two core word-learning indices to facilitate cross-study comparison (i.e., speeded recognition and semantic relatedness judgement). Experiment 1 (this chapter) adopted an experimental design similar to Johns et al.'s (2016) but considered in more detail topic familiarity. Simulation 1 (this chapter) is a computational study that used the reading materials developed for Experiment 1 to simulate spreading activation in lexical networks that differed in semantic diversity. Experiment 1 and Simulation 1 converged to show a low-diversity advantage in the learning of both orthographic forms and meanings, a finding attributed to what I refer to as "anchoring", a process of securing novel word representations onto long-term memory. Simulation 2 (Chapter 3) and Experiment 2 (Chapter 3) tested whether word learning might be better placed to take advantage of high semantic diversity if novel words were given an anchoring opportunity. Both the simulation and behavioral data showed that given an anchoring opportunity, novel word in the high-diversity condition were advantaged in visual recognition, standing in contrast to the results from Experiment 1 and Simulation 1. Simulation 3 (Chapter 4) used word vectors to mathematically simulate the learning of meanings across Experiments 1 and 2 and showed that words experienced in high-diversity environments tend to reside in more sparse semantic space. Experiment 3 (Chapter 5) adopted a highly comparable design as Experiment 2, but participants' eye movements were monitored during reading. The results suggested that high-diversity words were encoded more strongly during exposure and might have gained greater context-independence. Experiment 4 (Chapter 6), building on the notion of anchoring, asked whether providing readers with an explicit definition of the novel words could serve to anchor, and hence allow the words to be better placed to take advantage of contextual variation. The results are largely in line with those from the previous experiments and further demonstrated that explicit definitions could serve to anchor. Together, this research elucidates how variation in contexts, as indexed by semantic diversity, influences the learning of novel words during independent reading in young adults.

Experiment 1

The influence of semantic diversity on incidental word learning

2. Introduction

Following Johns et al. (2016), young adults read passages containing novel words; there was no instruction to learn the words as this experiment aimed to capture incidental learning. In one condition, passages were set around the same theme (low semantic diversity). Word learning in this condition was compared with learning following exposure to words in different-themed passages (high semantic diversity).

The current experiment differed from Johns et al. in a number of ways. First, I increased the familiarity of the topics included in the exposure phase. Johns et al. used relatively obscure topics, including passages about Strauss-Kahn (a French politician), Aundhati Roy (an Indian writer), Braybrooke (a theologian), Moore's Law, and Sunni Islam, whereas the current ones contained more familiar topics such as Brexit, David Bowie, and Donald Trump. This was motivated by pilot work indicating that topic familiarity may influence the effect of semantic diversity (Mak & Nation, 2019); I was also keen to motivate participants to read for interest and capture word learning via regular reading experience, rather than active study. Second, Johns et al. measured learning using speeded recognition and semantic relatedness judgement; I added a third measure, namely a pseudoword inferiority task.

As this task has not been used to index word learning before, it is useful to set out the rationale here. The pseudoword inferiority effect, built upon the traditional word superiority effect (Reicher-Wheeler task; Reicher, 1969; Wheeler, 1970), refers to the phenomenon that certain letters are recognized more poorly when they are seen in a pronounceable pseudoword (e.g., B**E**NANA) relative to a string of Xs (e.g., X**E**XXXX). To illustrate, consider the pseudoword BENANA, which differs from the familiar word BANANA in terms of the second letter. BENANA is shown on the computer screen for a short time (e.g., 50 ms) and participants are then asked to indicate whether it was an E or an A that they saw in the second letter position. Grainger and Jacobs (2005) reported that letter identification is poor in pseudowords such as BENANA: participants tended to choose A more than E, relative to the control condition when the target letter appeared in the same letter position but in a string of Xs (e.g., XEXXXX). This phenomenon, known as the pseudoword inferiority effect, is presumably due to top-down interference from familiar words (in this case, BANANA).

Extending this rationale to capture word learning, if the word form of a novel target has been well-established, it should induce a pseudoword inferiority effect. For example, if a participant experienced a novel word such as OSSANIC while reading, and it has gained word-like status (e.g., being well-specified), it should then bias letter recognition for a related pseudoword such as OSSA**M**IC in this letter identification task; that is, if OSSA**M**IC is shown briefly followed by a forced-choice decision of N or M at the fifth position, having learned OSSANIC should bias false recognition of N, relative to a baseline condition of XXXXMXX. If, however, OSSANIC had not been learned well during reading, less bias should be seen. In Experiment 1, I asked whether the pseudoword inferiority effect is sensitive to word learning as induced by variations in contextual history.

2.1. Method

2.1.1. Participants

Forty-five adults from the Oxford University participant pool (13 males; M_{age} = 22.53; SD_{age} = 4.68) took part in this experiment for either course credits or payment (£10/hour). All were native English speakers with normal or corrected-to-normal eyesight and none reported history of any developmental disorders. Five participants were excluded: two because their accuracy rates in the comprehension question were 2 SDs below the sample mean, and the other three because their accuracy rates in speeded recognition were 2 SDs below the sample mean. All analyses reported here, unless otherwise specified, are based on the remaining 40 participants.

2.1.2. Materials and procedure

The experiment comprised two phases, (i) an exposure phase where variation in a word's semantic diversity was induced and (ii) a surprise test phase that measured knowledge and processing of the newly learned words. Both phases were programmed and run using Gorilla (www.gorilla.sc), an internet-based platform. Participants completed both phases in a single 45-minute session held in a computer laboratory; up to five participants were tested at a time, well-spaced in the room.

(i) Exposure phase

Ten relatively rare English words were chosen as base words (e.g., *mendacious*, *predilection*; mean SUBTLEX frequency: 0.454/million; full list in Appendix 1A). Twelve passages were constructed for each base word: six sampled the same discourse topic and formed the low

semantic diversity condition; the other six spanned six distinct topics, forming the high semantic diversity condition. The topics were chosen to be familiar to an average undergraduate in the UK. Examples included Brexit, David Bowie, and Donald Trump. The passages were excerpted and edited from respectable non-fiction sources (e.g., The New York Times, The Guardian). Table 2.2 shows a sample passage.

Latent Semantic Analysis confirmed that the degree of semantic overlap between the passages was significantly higher in the low-diversity condition (M = 0.49, SD = 0.10) than in the high-diversity condition (M = 0.19, SD = 0.07), $t(10) = 3.39$, $p = .006$. Passages in the two conditions were matched for the number of words (M_{Low}= 125.6 vs M_{High} = 128.5), $t(71) = -0.85$, $p = .394$, and readability, as indexed by Flesch Reading Ease test, $t(71) = 0.40$, $p = .691$.

The next step was to replace the 10 base words with pronounceable pseudowords (e.g., OSSANIC, LONBUST, APPURATE). Two sets of ten pseudowords were chosen from the English Lexicon Project (Balota et al., 2007) and matched on a range of lexical properties, including word length, number of orthographic neighbors, and bigram frequency (for the full list, see Appendix 1B). Each participant encountered pseudowords from one of the two sets, and assignment to pseudoword set was randomized.

Each participant encountered ten pseudowords, five in the high-diversity condition and five in the low-diversity condition. In other words, semantic diversity was manipulated within-subject. Assignment to the semantic diversity condition was fully counterbalanced on the item level. Participants read each item six times with frequency matched exactly across diversity conditions. They were naïve to the purposes of the experiment and were told that they would be reading English passages from a standardized test designed for non-native speakers of English. Note that the register of the passages was relatively formal to align with the fact that in real-life, unknown or rare words are typically encountered in more formal or academic texts. The aim was to reduce the likelihood of participants deducing that the target words were fake replacements for existing words.

A reading trial began with a fixation cross of 600 ms followed by a passage. A Likert scale appeared below the passage after 15 seconds. The participants rated how comprehensible the passage was on a continuous scale of 1 (No understanding) to 7 (Perfect understanding). They had a maximum of 60 seconds to read and rate each passage. Reading time was recorded. After the rating was submitted (or after 60 seconds had elapsed), the passage and scale were replaced by two comprehension questions; these served to encourage reading for meaning. The question format was a mixture of multiple-choice, fill in

the blank, and true-false. They were straightforward and did not relate to the target words. Pilot testing confirmed that question difficulty was equivalent across diversity conditions.

Each participant read a total of 36 unique passages, divided equally into six blocks. Block order was randomized, as was the trial order within each block. In summary, each participant experienced five pseudowords in the low-diversity condition and five in the high-diversity condition. Regardless of the assigned condition, each pseudoword was seen a total of six times in the exposure phase.

Table 2.2

A sample passage in Experiment 1.

Base word	mendacious
Replaced by	ossanic [50% participants] *OR* chrabble [50% participants]
Passage	Though Trump is quick to deride unfavorable news stories as 'fake news', the administration owes its existence to the way mainstream media platforms can accommodate an avalanche of stories. Whether negative or fawning, any media attention popularize Trump's fact-free announcements and racist tweets. Through the sheer audacity of his ossanic personality, amplified a million-fold by his occupancy of the most powerful office on the globe, Trump is a one-man train-wreck. And, with each jarring twist and turn, the nation's media dutifully conveys his evermore crazy mutterings. While those in the mainstream media like to portray themselves as impartial observers, which at times forces them to take an antagonistic position towards the White House's shenanigans, more often than not reporters and editors are more like trumpets being played by a blow-hard president.
Note	Target words were **not** underlined in any of the passages that participants read.

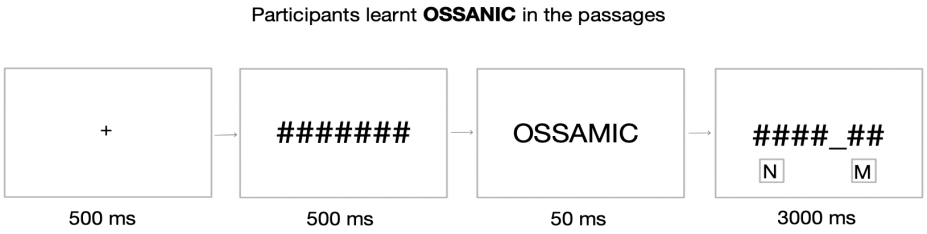
(ii) Test phase

Immediately following the exposure phase, participants completed three post-tests in a fixed order, namely pseudoword inferiority, speeded recognition, and semantic relatedness judgement.

Pseudoword inferiority. This was modelled on the Reicher-Wheeler task (see Figure 2.1). A trial started with a fixation cross for 500 ms, followed by a 500 ms mask. Next, a string of letters appeared for 50 ms. This was then replaced by a string of hashtags surrounding an underscored blank space. Participants were asked to select (using the mouse) which letter had appeared in the blank space, choosing from the two alternatives presented underneath the stimulus. The following trial started automatically once a response was recorded or after 3000 ms had elapsed.

Figure 2.1

A sample trial in the pseudoword inferiority task, Experiment 1.



There were three conditions in this task: novel word, X, and real word. For the novel word condition, stimuli were derived from the 10 newly learned pseudowords but deviating by one letter (e.g., OSSA**N**IC became OSSA**M**IC). The deviating letter was always the letter replaced by an underscore. In this example, the two alternative choices were M (the correct answer) and an N (incorrect answer, but aligning with the recently learned item). To increase power, there were two trials for each of the newly learned words, each with a different deviating letter (e.g., for OSSANIC, test stimuli were OSSA**M**IC and OSS**E**NIC). This increased the number of observations per diversity condition from five to 10 for each participant.

Participants also saw each deviating pseudoword in the X condition. For example, if OSSANIC was encountered during the exposure phase, they would see OSSAMIC and XXXMXX in this task, with M as the correct answer and N as the incorrect answer.

The real word condition comprised 18 medium-frequency words (e.g., FIREWORK; mean SUBTLEX frequency: 10.5/million). One letter was changed to make it a pronounceable pseudoword (e.g., FIREWORK → FIREWARK). Participants chose between the deviating letter (e.g. A, the correct answer) and the letter that would produce the real word (e.g. O, the incorrect answer). Items in the condition served as the positive control.

There were a total of 56 experimental trials, whose order was randomized. The task started with five practice trials, where familiar words (e.g., DESKTOP, LAMP) were shown.

***Speeded recognition*.** Modelled on a standard lexical decision paradigm, participants saw a pseudoword centered on the screen and were asked to determine whether they had encountered that in the reading passages, responding as quickly and as accurately as possible.

Following two practice trials, each trial began with a fixation cross displayed for 800 ms, followed by a stimulus for 1800 ms, during which the participants pressed buttons to make a judgement. For half of the participants, the A key was used for “Yes: I have encountered this word” and the L key for “No: I have not encountered this word”; the

assignment of keys was reversed for the other participants. The next trial started as soon as a response was recorded, or if no response was recorded after 1800 ms. No feedback was given. Both accuracy and RTs were recorded.

Of the 20 pseudowords selected for the experiment (see exposure phase), each participant saw 10 of them in the reading passages; these were presented for speeded recognition, along with the other 10 unstudied pseudowords, which served as foils here. The 20 items were presented in a random order in a single block. The block was repeated three times in total with a random order each time. Overall, each participant completed 60 trials: 30 'yes' items (the 10 newly learned pseudowords x 3) and 30 'no' items (the 10 foils x 3).

Semantic relatedness. In each trial, participants saw one of the 10 newly learned pseudowords on the left side of the computer screen (e.g., OSSANIC) and a familiar English word (e.g., DISHONEST) on the right side. Their task was to judge, as quickly and as accurately as possible, whether the two words were related in meaning. Each pseudoword was seen on four occasions, twice paired with an associated word ('related' trials) and twice with an unrelated word ('unrelated' trials).

To generate the associated words, three native speakers of English who did not participate in the main study were asked to give three semantic associates for each of the 10 base words (e.g., *mendacious*). From these responses, the two most popular ones were selected, none of which appeared in the passages (e.g., for *mendacious*, the associates were *dishonest* and *fraudulent*; see Appendix 1C for the complete list). To create the unrelated trials, each pseudoword was paired with two different words unrelated in meaning, both randomly taken from the list of associates.

Each trial started with a fixation cross for 500 ms, then two words appeared on the screen. Participants had 2500 ms to decide if they were related in meaning. Stimuli were presented in black, uppercase courier-new font on a white background. For half of the participants, the left arrow key was used for "They are related" and the right arrow key for "They are unrelated"; the assignment of keys was reversed for the other participants. After a response, or if no response was recorded after 2500 ms, the next trial started automatically. Both accuracy and RTs were recorded. Each participant completed 40 trials in total, divided into two blocks. Block order was randomized, as was the trial order within each block. This task started with a practice block containing 10 trials with familiar English words as stimuli (e.g., EYE——EAR).

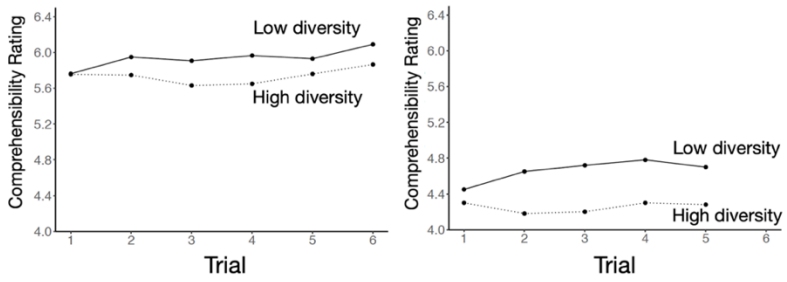
2.2. Results

2.2.1. Exposure phase

Performance on the comprehension questions was satisfactory (*M* = 84.5%, *SD* = 5.96%) indicating that the passages were read for meaning. Comprehension was equivalent across the semantic diversity conditions, *t*(39) = 0.12, *p* = .76. There was however a diversity effect on passage reading time, with passages in the high-diversity condition (*M* = 36.05 sec, *SD* = 10.3 sec) taking longer to read than those in the low-diversity condition (*M* = 34.79 sec, *SD* = 9.9 sec), *t*(39) = 3.23, *p* = .002. While all passages received a high comprehensibility rating, those in the low-diversity condition were rated as more comprehensible (*M* = 5.94, *SD* = 0.96) than those in the high-diversity condition (*M* = 5.74, *SD* = 0.99), *t*(39) = 4.58, *p* = < .001. This suggests that subjective comprehension of the low-diversity passages (all on the same topic) grew as participants became more familiar with that topic. Note Johns et al. (2016) reported an average ratings of about 4.5 (out of 7); as intended, therefore, the current passages were more comprehensible (see Figure 2.2 for comparison), probably due to increased topic familiarity.

Figure 2.2

Comprehensibility ratings for high and low diversity passages across trials in Experiment 1 (left) and Johns et al. (right). Higher ratings (max = 7) indicate that passages in Experiment 1 were easier to comprehend.



2.2.2. Test phase

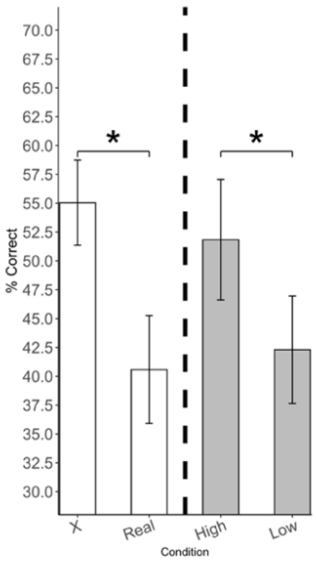
Analysis approach. All data were analyzed in a mixed-effects environment using the lme4 package (version 1.1.17; Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2019). Generalized linear mixed-effect models were used for accuracy data while RT data were analyzed using linear mixed-effect models. RT data were transformed using either inverse or log transformation; the one that resulted in more normal residual distribution, as indexed by qqPlot (Millard, 2013), was adopted. All the models included random effects for participants and items and a fixed effect for semantic diversity, which had two levels (low

diversity vs. high diversity). This was coded using sum contrasts. Following Barr et al. (2013), models were first computed with a maximal random structure. If a model failed to converge or produced a singular fit, the random-effect structure was simplified (e.g., Barr et al., 2013; Bolker et al., 2009). The explanatory value of the converged model was compared with a random-intercept-only model, using anova model comparison. If the two models did not differ (as indexed by a threshold of chisq < 3; Jaeger, 2011; see also Gaskell, Cairney, & Rodd, 2019), the random-intercept-only model was favored as this is considered more statistically powerful (Matuschek et al., 2017). However, if the two models differed (chisq > 3), the conservative model with the maximal (or the near-maximal) structure was favored and reported (all data and scripts are available on OSF).

Pseudoword inferiority. Accuracy across the four conditions is plotted in Figure 2.3. The first point to highlight is that the basic pseudoword inferiority effect was present, replicating earlier findings (Grainger & Jacobs, 2005): Divergent letters embedded in real words (e.g., the A in FIREWARK) were recognized more poorly (*M* = 40.5%) than letters embedded in a string of Xs (*M* = 55%), *z* = 2.43, *p* = .015.

Figure 2.3

Mean accuracy rates across the four conditions in the pseudoword inferiority task, Experiment 1.



Note. (i) Lower performance in the real word and low conditions is indicative of greater top-down interference. (ii) Error bars represent 95% within-subject confidence intervals.

Turning to the newly learned words experienced in the exposure phase (gray bars in Figure 2.3), letter identification was significantly poorer in the low-diversity condition (*M* = 42.8%) relative to the high-diversity condition (*M* = 51.8%; *z* = -2.20, *p* = .028; see Table 2.3 for model summary). This suggests that the low-diversity passages enabled the novel words to establish stronger and/or more well-specified lexical representations, leading to greater error in this letter identification task.

Table 2.3

*Summary of the GLME models examining the effect of semantic diversity on **accuracy** in each of the post-tests, Experiment 1.*

	Estimate	Standard Error	z value	p value
(a) Pseudoword inferiority				
Intercept	-0.20	0.21	-0.98	.325
SemD	0.18	0.08	2.20	.028*
(b) Speeded recognition				
Intercept	3.48	0.42	8.30	<.001*
SemD	0.67	0.31	2.12	.034*
(c) Semantic relatedness				
Intercept	0.66	0.12	5.48	<.001*
SemD	0.40	0.06	6.60	<.001*

Table 2.4

*Summary of the LME models examining the effect of semantic diversity on **RTs** in each of the post-tests, Experiment 1.*

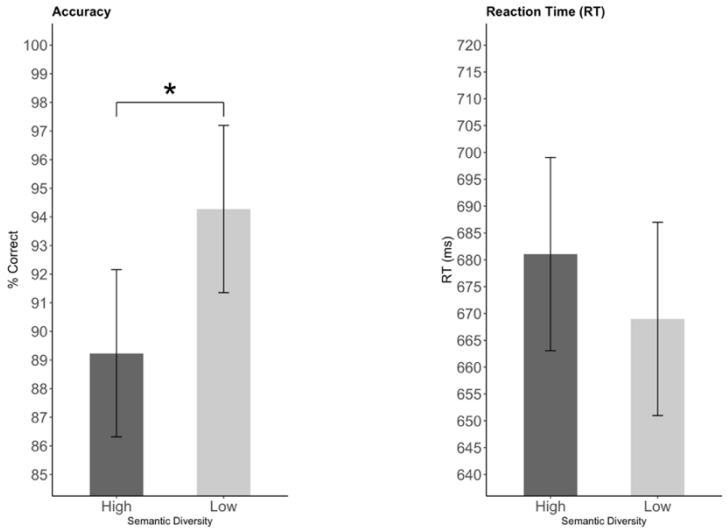
	Estimate	Standard Error	t value	p value
(a) Speeded recognition				
Intercept	-155.9	2.75	-56.77	<.001*
SemD	0.62	0.87	0.71	.477
(b) Semantic relatedness				
Intercept	721.1	2.30	313.33	<.001*
SemD	1.91	0.81	2.36	.019*

Note. To improve interpretability, estimate and standard error were multiplied by 100,000 in speeded recognition and by 100 in semantic relatedness.

Speeded recognition. Data from two participants were lost due to equipment failure, meaning these analyses are based on 38 participants. Only "yes" trials were analyzed. Mean accuracy rates in the two diversity conditions are summarized in the left panel of Figure 2.4. Accuracy was significantly higher in the low-diversity condition (*M* = 94.4%) than in the high-diversity condition (*M* = 89.3%; *z* = 2.12, *p* = .034; see Table 2.3 for model summary). Turning to RT, only correct responses were analyzed (91.8% of all "yes" trials). One trial was removed as it was more than 2.5 SDs below the mean RT. Pseudowords experienced in the low-diversity condition elicited faster response (right panel of Figure 2.4), but this difference was small at 10 ms, and not statistically significant (*t* = -0.712, *p* = .477; see Table 2.4 for model summary).

Figure 2.4

Mean accuracy (left) and RTs (right) in the two diversity conditions in speeded recognition, Experiment 1.



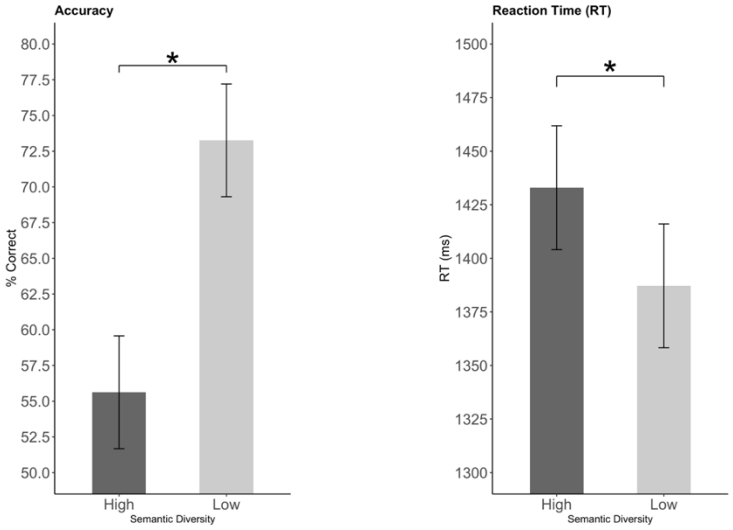
Note. Error bars represent 95% within-subject confidence intervals.

Semantic relatedness. Overall accuracy was satisfactory (*M* = 64.5%, *SD* = 11.4%) and was above chance [one-sample t-test: $t(39) = 8.01$, $p < .001$], indicating that participants had gained some knowledge of the meanings of the novel words. The accuracy data are summarized by diversity condition in Figure 2.5, and model output is shown in Table 2.3. Performance was significantly more accurate in the low-diversity condition, *z* = 6.60, *p* < .001. RTs to correct trials (14.2% error trials were excluded) patterned in the same

direction with faster judgement to novel words encountered in the low-diversity passages (t = 2.36, p = .019; see Table 2.4 for model summary).

Figure 2.5

Mean accuracy (left) and RTs (right) in semantic relatedness judgement for words experienced in the two diversity conditions, Experiment 1.



Note. Error bars represent 95% within-subject confidence intervals.

2.3. Discussion

In Experiment 1, I induced variation in a novel word's semantic diversity at exposure and investigated whether this influenced the learning and processing of the novel words. Young adults read passages in which pseudowords (e.g., *ossanic*) replaced low-frequency words (e.g., *mendacious*). For each participant, five novel words appeared in passages built around the same discourse topic (low semantic diversity) while another five novel words appeared in passages spanning six distinct discourse topics (high semantic diversity). Word learning was assessed using three different measures, namely pseudoword inferiority, speeded recognition, and semantic relatedness judgement. Across all three tasks, novel words experienced in the low-diversity condition were advantaged relative to the same words seen in the more diverse condition. Specifically, the finding that low-diversity novel words were recognized more accurately in speeded recognition contrasts with Johns et al. (2016), who reported more accurate recognition for novel words experienced in *more* diverse contexts. Given the experimental approach is so similar, why should the results pattern in an opposite direction?

One possibility is that the reading materials in the current experiment centered around familiar themes whereas Johns et al. (2016) used more obscure material. Consistent with this, the passages in Experiment 1 were rated as more comprehensible than those used by Johns et al. (see Figure 2.2 for comparison). It is well-established that when a memory target—be it a novel word or a novel image—is encountered, it is encoded with reference to what is already known (e.g., Nelson & Shiffrin 2013). Novel items that are congruent with pre-existing knowledge are more likely to be remembered than those low in congruence (e.g., Greve, Cooper, Tibon, & Henson, 2019; Kaefer et al., 2015; Kan, Alexander, & Verfaellie, 2009). For word learning specifically, Pulido (2004) asked adult L2 learners of Spanish to read four Spanish texts, two describing familiar situations and two describing less familiar scenarios. Target words encountered in texts describing the more familiar scenarios were better retained, both immediately and 28 days later (see also Diakidoy, 1998; Nagy et al., 1985). Overall, there is converging evidence that incidental word learning, at least in its initial stages, fares better when the discourse topic is more familiar (see section 2.3 in Chapter 1 for a review).

There are multiple explanations as to why greater topic familiarity is beneficial. For one, familiarity with the topic might help people infer the meaning of new words more easily (Freebody & Anderson, 1983). For another, when background knowledge is in place, comprehension resources can be allocated to processing the novel words, facilitating word learning (e.g., Cervetti, Wright, & Hwang, 2016; Chiesi et al., 1979). A different type of explanation builds from Steyvers and Tenenbaum (2005). When new words are introduced into the mental lexicon, they build associations with pre-existing words and knowledge. When topic familiarity is high, new words have greater opportunity to “anchor” to the topic and its pre-existing representations in long-term memory (Mak, 2019). Secure anchorage with long-term memory might help word learning and retention by stabilizing the word representation and reducing interference (Mak, 2019; Wang et al., 2017). In contrast, when the topic is less familiar, novel words are less likely to “anchor”, resulting in more fragile memory representations of the words.

Topic familiarity was high across both diversity conditions in my experiment, yet more learning was seen for words experienced in the low-diversity condition. A similar effect has also been seen in word learning experiments with toddlers and children (e.g., Horst et al., 2011; Tamura, 2015; Williams & Horst, 2014). One explanation, following Horst’s (2013) work with children, is that more similar contexts help channel attention from comprehending the texts to processing the novel words, thereby boosting word learning. It

is not clear that this explanation best fits the low-diversity advantage seen in Experiment 1. While the discourse topic remained the same in the low-diversity condition, the passages actually varied from episode to episode across both conditions (i.e., not repeated verbatim); note also performance on the comprehension questions was high and equivalent across conditions, and the discourse topics were familiar to participants. Arguably, the two diversity conditions differed in the opportunity they offered for new words to anchor onto pre-existing knowledge. On this view, the low-diversity passages (i.e., six encounters in the same discourse topic) provided more opportunity for secure anchorage than the high-diversity passages (i.e., six encounters in six different discourse contexts). This way, the nascent representations of the low-diversity words may be stabilized and reinforced to a larger extent. A consequence of this may be greater accessibility in memory. Future work is needed to examine what kind of prior knowledge facilitates anchoring (e.g., Mak & Twitchell, 2020) and how this anchoring process may be influenced by the amount of prior knowledge accumulated over the life span. For instance, Ramscar, Sun, Hendrix, and Baayen (2017) reported that in older adults, accumulated associations that an existing word has impeded the formation of new associations, suggesting the possibility that older adults may not benefit from anchoring as much as young adults (see also Qiu and Johns, 2020).

Returning to the question why the current results differed to those obtained by Johns et al. (2016), high semantic diversity, as Johns et al. argued, may boost learning by encouraging readers to allocate more attentional resources to encoding the novel words. This is supported by their observation that reading times were significantly longer for high-diversity passages. In Experiment 1, I also observed the same difference in reading times; despite this, a low-diversity advantage emerged in the post-tests, suggesting that the attention-related benefit associated with high-diversity environments was offset perhaps by topic familiarity, hence bringing about the different pattern of results.

A key limitation of Experiment 1 is that it is unclear the extent to which the observed low-diversity advantage can be attributed to a third factor such as arousal and differences in pre-existing knowledge and attitudes. For instance, the pseudoword *ossanic*, based on *mendacious*, appeared in six different passages centered on Donald Trump in the low-diversity condition; in the high-diversity condition, *ossanic* was embedded in six different-themed passages: one about North Korea, one about Nigel Farage, and one about childhood memory, etc. It is reasonable to argue that the six Donald Trump passages were more similar in terms of their level of arousal and the amount of background knowledge required. In other words, the low-diversity advantage in Experiment 1 can be attributed, at least

partially, to these differences instead of semantic diversity *per se*. In light of this, I conducted a computational study where spreading activation was simulated in lexical networks differing in semantic diversity.

Simulation 1

Simulating spreading activation in lexical networks that differ in semantic diversity (1)

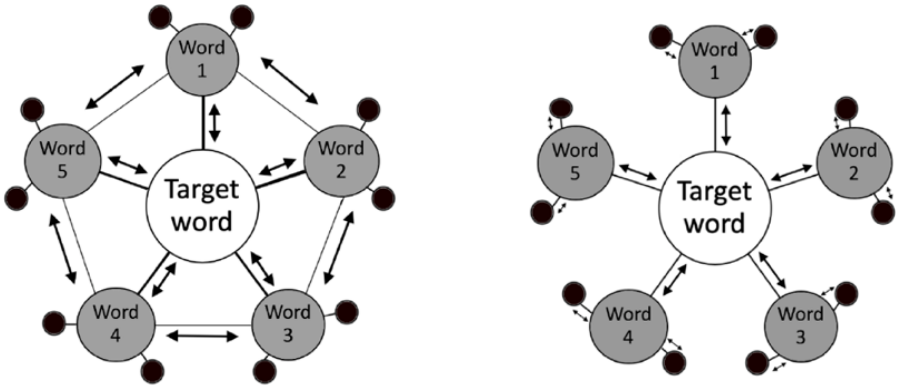
3. Introduction

Here, I used the reading materials developed for Experiment 1 to investigate whether target words appearing in the low-diversity passages would show more efficient processing as estimated by a computational simulation. This offered an opportunity to replicate the behavioral findings, and to see whether contextual variation influences how words are learned and processed when other factors such as arousal become irrelevant. Before going into the details of the simulation, I first consider the background from which the simulations were derived.

According to localist approaches, each word in the mental lexicon is represented by a distinct node in a network, and words are connected if they are phonologically, orthographically, and/or semantically related (e.g., Collins & Loftus, 1975; Mak et al., 2020; Siew & Vitevitch, 2020). An accumulating body of evidence suggests that where the word stands in a network influences how efficiently its form is processed (e.g., Chan & Vitevitch, 2009, 2010; Siew & Vitevitch, 2016). Of particular interest here is a network structure known as clustering coefficient. This refers to the degree to which neighbors of a word node are also neighbors of each other. In a word learning experiment, Goldstein and Vitevitch (2014) found that adult participants retained more novel words with higher clustering coefficients. They argued that for such words, the greater interconnections among the immediate neighbors (see left panel of Figure 2.6) help activations circulate back to the novel word node (note: activation is assumed to spread bilaterally), thereby strengthening its nascent representation. In contrast, for novel words with a lower clustering coefficient (see right panel of Figure 2.6), activation is more likely to disperse outside of the immediate neighborhood, reducing the amount of activation circulating back to the novel word target. In turn, this lowers its likelihood of being retained.

Figure 2.6

Left: *A target word with a higher clustering coefficient.* ***Right:*** *A target word with a lower clustering coefficient.*



Given Goldstein and Vitevitch's (2014) finding, I hypothesized that the low semantic diversity condition in Experiment 1 led to higher clustering coefficients for the novel words (relative to those in the high-diversity condition), thereby boosting the amount of activations they received and leading to better learning. To test this idea, I first computed term-term co-occurrence matrices for each target word, using the reading materials from Experiment 1. Two sets of networks were then constructed for each word, one low-diversity and one high-diversity. Here, the networks are assumed to reflect the target word's contextual history, which in turn exerts an influence on the efficiency with which the word is identified in tasks such as speeded recognition. Next, spreading activation was simulated across the networks, allowing me to see if target words appearing in the low-diversity network receives more activation. I used "spreadr", a package that has simulated behavioral data from psycholinguistic tasks that tap single-word recognition (Siew, 2019). It is unclear whether the package can be extended to more complex tasks such as semantic judgement. Therefore, the simulation data are most appropriately considered against the behavioral data from the speeded recognition task in Experiment 1.

Like all computational simulations, some assumptions are necessary. Concerning topic familiarity, it is assumed that if a target word is experienced in familiar topics, it will be connected to well-established nodes that send activations back, thereby reinforcing the target's representation. In contrast, if a target word is experienced in unfamiliar topics, it is assumed that the target will be connected to impoverished or previously non-existent nodes, resulting in less activation being sent back. In the present simulation, activations in all the networks were allowed to flow relatively freely; this is to reflect that all the passages in

Experiment 1 were based on familiar topics, which could presumably reinforce the targets by sending back activations. If one wishes to simulate the effect of *unfamiliar* topics, the free flow of activations could be restricted so that the target node receives fewer reinforcing activations from its neighboring nodes. The second assumption is that activations in both the high- and low-diversity networks were allowed to spread at the same rate, meaning that every topic is set to be "equally capable" at sending activations back to the target. This way, the levels of arousal and required background knowledge would be computationally matched between the two diversity conditions.

3.1. Method

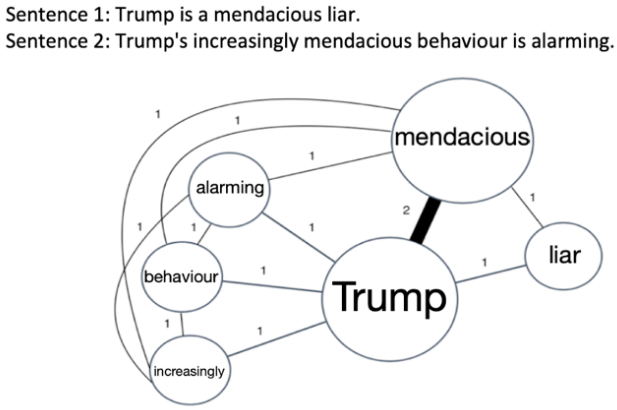
I used eight of the 10 base words from Experiment 1 for this simulation (*mendacious*, *predilection*, *constellation*, *proliferation*, *diffident*, *latent, vagarious,* and *castigate*). [4] Two corpora were created for each word, one for the low-diversity passages and one for the high-diversity passages, totaling eight for each diversity condition. Punctuation, numerals, and function words (e.g., prepositions, pronouns) were removed. Proper nouns containing two words (e.g., New York) were merged (e.g., NewYork), and the corpora were stemmed using the "SnowballC" package (Bouchet-Valat, 2019) in R. Next, a global term-term co-occurrence matrix was created for each corpus, using the "chinese.misc" package (Wu, 2019) on the basis that co-occurrence provides "an objective measure that aggregates semantic, conceptual and syntactic relatedness" (Beckage, Smith, & Hills, 2011, p. 2). Unlike some previous studies which captured co-occurrence within a 5 to 10 word window around a target word (e.g., Beckage et al., 2011; Hills, Maouene, Riordan, Smith, 2010; Li, Burgess, & Lund, 2000; McDonald & Ramscar, 2001), I opted for global co-occurrence computed across passages. The reasoning was threefold. First, corpus size for each word was small (~400 words). Second, global co-occurrence provided an opportunity to capture the topic of each passage. For example, with a window size smaller than 10, the target word *mendacious* would not have "co-occurred" with *Trump* in most of the passages. Third, the passage was the "unit" used in Experiment 1, and was therefore retained here, rather than determining an arbitrary fixed window.

Sixteen co-occurrence matrices were created and transformed into a weighted, undirected network using the "igraph" package (Csardi & Nepusz, 2006). Figure 2.7 depicts a sample network created for two illustrative sentences.

[4] Two of the 10 target words (i.e., *avidity*, *shenanigans*) in Experiment 1 shared the same contextual environments with two other target words (i.e., *diffident*, *mendacious*). To avoid duplications, *avidity* and *shenanigans* were excluded.

Figure 2.7

A sample weighted undirected network.



Note. The number next to each link represents how many times two words co-occurred together in the corpus. This is also the *weight* of connection.

Spreading activation was then implemented in each network using the "spreadr" package (Siew, 2019). There are several parameters that can be specified in the simulation; full details of settings are available in Appendix 1D. Of note here is *retention*: This refers to the proportion of activation that remains in a node at each time step. For example, a value of 0.8 means that at each time step, 80% of activation that a node receives will be retained in the node while the remaining 20% will spread to the neighboring nodes. Following Siew (2019) and Vitevitch et al. (2011), I used nine retention values ranging from 0.1 to 0.9, in increments of 0.1. This was to ensure that the outcome of the simulations is generalizable across retention levels (Siew, 2019). Spreading activation within a network produced an activation value for each node at every time step. Following Siew (2019, p. 916), "the activation value of the target node at the final time step was taken as a proxy for the efficiency with which the word was retrieved from the lexicon. Specifically, higher activations [at the final time step] correspond to faster RTs and/or higher accuracy rates in psycholinguistic tasks". To summarize, I ran 144 simulations (8 words x 2 diversity conditions x 9 retention levels), 72 as high semantic diversity networks and 72 as low diversity networks.

3.2. Simulation results

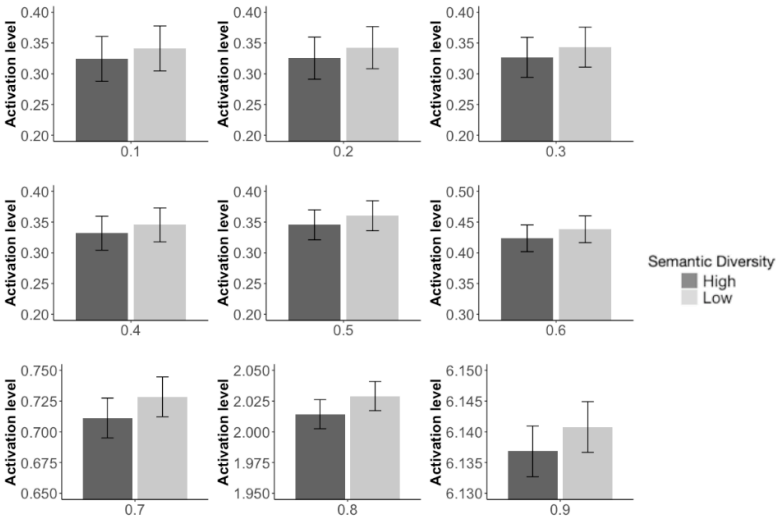
Figure 2.8 shows the average activation levels in the two types of diversity networks across the nine retention levels. Mean activation of the target words was consistently higher in the low-diversity networks than in the high-diversity networks across all the retention levels. To

test whether the difference is statistically significant, I used a paired Wilcoxon signed-rank test, as the distribution was non-normal. It confirmed that activation levels of the target words were reliably higher in the low-diversity networks (M = 1.230) than in the high-diversity networks (M = 1.216), $Z(143) = 2.66$, $p = .004$.

Next, I obtained the clustering coefficients of the target nodes in the two types of diversity networks. As shown in Figure 2.9, clustering coefficients of the target nodes were consistently higher in the low-diversity networks (M_{High} = 0.206 vs. M_{Low} = 0.249), $t(7) = -6.06$, $p < .001$.

Figure 2.8

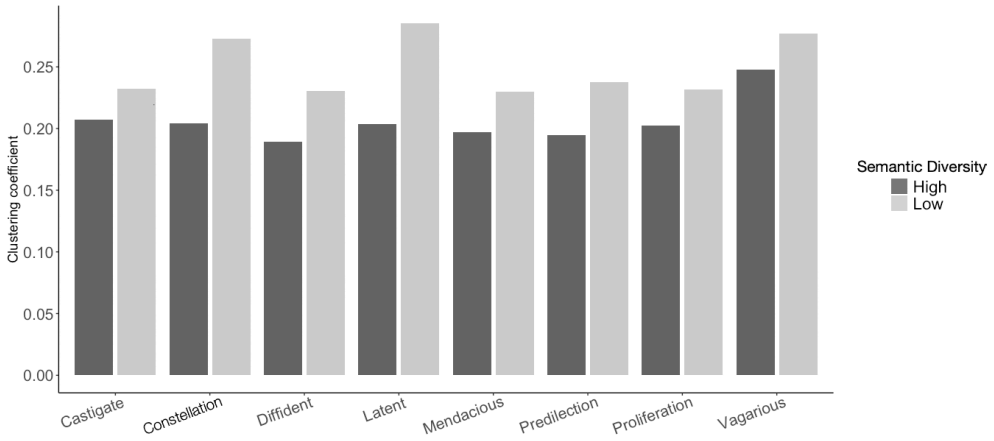
Mean final activation values in the target node in each type of diversity networks across the nine retention levels in Simulation 1.



Note. Error bars represent 95% within-item confidence intervals.

Figure 2.9

Clustering coefficients of the target nodes in the two types of diversity networks in Simulation 1.



3.3. Discussion

Simulation 1 provided computational evidence that in the initial stages of word learning, words experienced in low-diversity (but topically familiar) contexts are advantaged in visual recognition relative to the same number of encounters in more diverse contexts. This processing advantage aligns with the behavioral data in Experiment 1, where words experienced in the low-diversity condition were recognized with greater accuracy. Simulation 1 eliminated some of the potentially confounding variables from Experiment 1, including for example arousal and the different levels of required background knowledge. The emergence of a low-diversity advantage in the simulation data therefore supports the argument that low-diversity contexts are more conducive to the initial stages of incidental word learning.

The target nodes received more activation when they appeared in the low-diversity networks, plausibly because these nodes had higher clustering coefficients, reflecting a more dense immediate neighborhood. This interpretation stems from Goldstein and Vitevitch (2014) who found that novel word forms with higher clustering coefficients were better recalled in their word-learning study. One benefit of having a high clustering coefficient, according to the authors, is that it helps activation to circulate back to the target nodes, which in turn strengthens their nascent representations. In short, situating new words in more close-knit (and topically familiar) contexts may boost initial learning. This description also captures the notion of anchoring as discussed in Experiment 1 in that low-diversity

passages allowed the novel words to be secured onto long-term knowledge, thereby boosting word learning and retention.

Finally, it is important to stress that the current simulations have at least two limitations. First, they were built upon assumptions derived from spreading activation dynamics (Anderson, 1983; Collins & Loftus, 1975). For instance, it was assumed that a word is represented as a single node in the lexicon. While the simulations serve as an existence proof of the behavioral findings, they do not rule out other possible accounts based on other assumptions. For instance, it is possible that the present behavioral findings could fit with a connectionist framework (e.g., McClelland & Rumelhart, 1981; Harm & Seidenberg, 2004; Plunkett, 1998), where a word is not represented as a single node but as distributed activity patterns. Future work is needed to explore how well these behavioral findings fit with other learning/memory frameworks. Second, the networks constructed were based solely on the reading materials I developed; for instance, the passages with *mendacious* as the target word had words like Trump, White House, tweets, and Republican, so these words were also present in the networks. These words represent (presumably) familiar concepts, so in the simulation, they were set to be capable of reinforcing the target node by sending back activations. In other words, the networks did have some prior knowledge (e.g., Trump, Republican) built into them. Clearly, however, the amount is limited and is some distance from human learning, where word learning is influenced by a large and complex base of pre-existing knowledge (e.g., Ramscar et al., 2017).

In summary, Experiment 1 and Simulation 1 both suggested that the initial stages of word learning can profit from words being experienced in familiar passages that overlap in semantic content, relative to the same amount of experience in familiar passages that are contextually diverse. Importantly however, this low-diversity learning advantage is at odds with what is known about established words where there is ample existing evidence to show that high semantic diversity words are identified more efficiently, even when frequency and other key lexical statistics are controlled for (e.g., Hoffman & Woollams, 2015; Hsiao et al., 2020; Jones et al., 2012; Pagán et al., 2020). The contrasting results between these effects on established words and the current findings for newly learned words suggests that there may be a switch in direction in how contextual variation influences visual word recognition. With this as a backdrop, I explored in the next chapter how a low-diversity advantage in the initial stage of word learning becomes a high-diversity advantage as word knowledge matures.

Chapter 3

Data from Experiment 1 and Simulation 1 converged to show a recognition advantage for novel words experienced in contexts that are *less* semantically diverse. However, given there is good evidence from corpus-based studies that known words higher in semantic diversity are identified more accurately and rapidly in lexical decision (e.g., Hoffman & Woollams, 2015; Hsiao & Nation, 2018; Jones et al., 2012), I explored in this chapter how does a high-diversity advantage come to be. This chapter comprises a network simulation (Simulation 2) and a behavioral study (Experiment 2).

Simulation 2

Simulating spreading activation in lexical networks that differ in semantic diversity (2)

1. Introduction

I hypothesized that if new words are securely anchored to a familiar topic, word learning might be better placed to take advantage of contextual variation. In Simulation 2, the same simulation method was used as before, but all target words were embedded in one discourse topic multiple times before contextual variation was induced at a second step. This design is best explained with reference to Table 3.1.

Table 3.1

How semantic diversity was implemented in Simulation 2 (and Experiment 2).

	Exposure phase							
	Anchoring passages					Post-anchoring passages		
	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Passage 6 (Each contained 2 target words)	Passage 7 (Each contained 2 target words)	Passage 8 (Each contained 2 target words)
High-diversity word	Topic A	Topic A	Topic A	Topic A	Topic A	Topic B	Topic B	Topic B
Low-diversity word	Topic B	Topic B	Topic B	Topic B	Topic B			

Passages 1 to 5 constituted the anchoring phase. Here, all word targets, regardless of the diversity condition they were assigned to, appeared in passages centered around one discourse topic. After anchoring, the diversity manipulation was introduced. Target words in the high-diversity condition appeared in a different topic in passages 6 to 8. In contrast, target words in the low-diversity condition remained within the same topic throughout all

eight passages. With some anchorage in place, I predicted that target words in the high-diversity condition would show a processing advantage.

Note that, the number of exposure per word was eight here, not six as in Experiment/Simulation 1. I purposely did not match the number of exposure as it was hypothesized that a high-diversity recognition advantage would emerge when novel words were securely anchored. Experiment/Simulation 1 provided evidence of secure anchorage following six exposure. To test my hypothesis, therefore, it was necessary to increase the number of exposure beyond six.

1.1. Materials

The materials for Simulation 2 were designed to be used in a subsequent behavioral experiment, reported later in this chapter. Therefore, the passages were constructed to be appropriate for young adults to read in approximately 30 to 45 minutes, allowing me to maintain the single exposure session used in Experiment 1. All passages had high topic familiarity.

I used the same 10 base words as in Experiment 1 (e.g., *mendacious*) plus two additional words (e.g., *architecture*, *imbroglio*) to increase power. As before, two sets of passages were created for each target word: The low-diversity set comprised eight passages; their discourse topic was the same (e.g., Donald Trump). The high-diversity set comprised five passages on one topic (e.g., a Michael Jackson documentary) and three passages on another topic (e.g., Donald Trump).

Each target word, regardless of its assigned diversity condition, appeared in one discourse topic for passages 1 to 5 (see Table 3.1). These passages are referred to as the "anchoring passages". A total of 120 anchoring passages were created (= 12 target words x 5 exposure x 2 diversity), each containing one target word and an average of 36.2 words. The high and low diversity conditions were matched on the number of words [M_{High} = 37.3 vs. M_{Low} = 35.1; $t(119) = 1.62$, $p = .11$] and readability, as indexed by Flesch Reading Ease test [M_{High} = 41.2 vs. M_{Low}= 45.1; $t(119) = -0.71$, $p = .49$].

In addition, 18 passages from Experiment 1 were chosen and modified to serve as the post-anchoring passages (passages 6 to 8, see Table 3.1). Each contained two target words, one belonging to the high-diversity condition and one to the low-diversity condition.

1.2. Simulation procedure

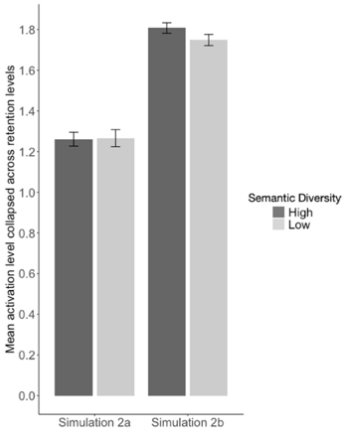
Spreading activation was simulated in broadly the same way as Simulation 1, using the same procedures and packages. Simulation 2a used only the anchoring passages as input (i.e. passages 1-5). Since the anchoring passages were "identical" across the two diversity conditions, no difference in activation values was expected. Simulation 2b included the post-anchoring passages as well as the anchoring passages. If post-anchoring semantic diversity influences the structure of the networks, there should be a difference in activation levels between the two conditions in Simulation 2b. Of interest is how this pattern compared to Simulation 1 which did not contain an anchoring phase. I hypothesized that adding diversity post-anchoring should result in the target words in the high-diversity networks receiving more activation than those in the low-diversity networks, in contrast to Simulation 1. Overall, I ran a total of 432 simulations, half for Simulation 2a, the other half for Simulation 2b. For each set of simulation, 108 were on high-diversity networks and 108 on low-diversity networks, each comprising 12 words x 9 retention levels.

1.3. Simulation results

Figure 3.1 shows the mean activation values of the target nodes, collapsed across retention levels, between the high and low diversity networks in both simulations. Simulation 2a included only the anchoring passages. As expected, and as shown in Figure 3.1, there was no difference in activation values between target nodes appearing in the high (M = 1.261) and in the low (M = 1.265) diversity networks. This non-difference was confirmed in a Wilcoxon signed-rank test, $Z(107) = -0.55$, $p = .292$.

Figure 3.1

Mean activation values of the target nodes, collapsed across retention levels, between the high and low diversity networks in Simulations 2a and 2b.

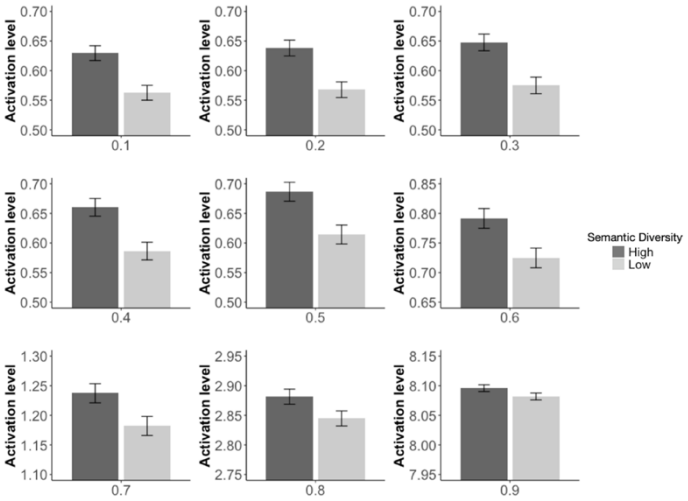


Note. Error bars represent 95% within-item confidence intervals.

In contrast, Simulation 2b, which included the post-anchoring passages into the networks, produced higher mean activation levels in the high-diversity networks (M = 1.807) than in the low-diversity networks (M = 1.748). A Wilcoxon signed-rank test confirmed that target nodes in the high-diversity networks received significantly more activations than those in the low-diversity networks, $Z(107) = -8.90$, $p < .001$. This difference was also evident across the nine retention levels, shown in Figure 3.2.

Figure 3.2

Mean final activation values in the target node in each type of diversity networks across the nine retention levels in Simulation 2b.

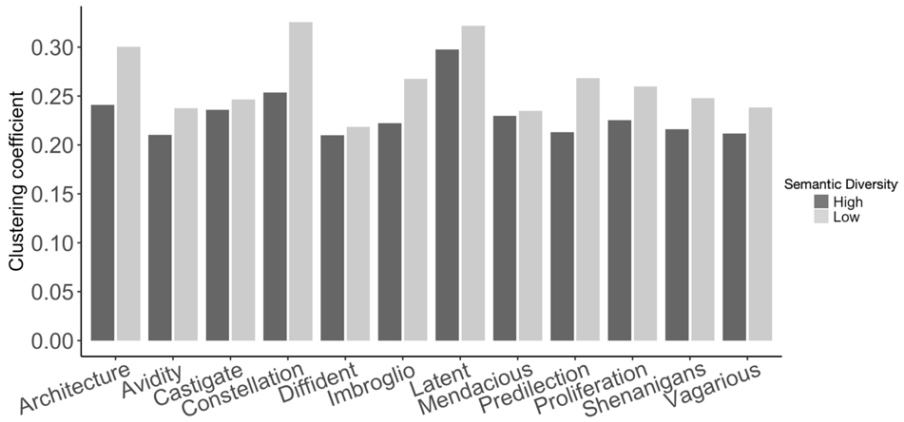


Note. Error bars represent 95% within-item confidence intervals.

Finally, I checked the clustering coefficients of the target nodes (i.e., the probability with which the neighbors of the target are also neighbors of each other) in the two types of diversity networks. As in Simulation 1, clustering coefficients of the target nodes were consistently higher in the low-diversity condition (M_{High} = 0.231 vs. M_{Low} = 0.264; see Figure 3.3), $t(11) = -3.30$, $p < .001$.

Figure 3.3

Clustering coefficients of the base words in the two types of diversity networks in Simulation 2b.



1.4. Discussion

Simulation 2 tested the hypothesis that if new words are securely anchored to a familiar topic, word learning might be better placed to take advantage of contextual variation. In Simulation 2a, there was no difference in the amount of activation for target words in the two types of diversity networks, as expected. However, in Simulation 2b, adding diversity via the post-anchoring passages changed the pattern of results such that target words in the high-diversity networks received reliably more activations. Previous work has attributed activation level to efficiency of lexical processing (e.g., Siew, 2019). The results of Simulation 2b are therefore consistent with corpus-based studies in showing a recognition advantage for words higher in semantic diversity (e.g., Hoffmann & Woollams, 2015; Hsiao & Nation, 2018; Jones et al., 2012).

Then, I considered why the novel words appearing in the high-diversity networks received more activations in Simulation 2b. In Simulation 1, the low-diversity advantage was attributed to higher clustering coefficients. Clearly, this cannot explain the high-diversity advantage seen in Simulation 2b, as the clustering coefficients were also always lower in the high-diversity condition (M_{High} = 0.231 vs. M_{Low} = 0.264). Instead, increased activation in the

high-diversity condition may stem from the target words being associated with one additional discourse topic, meaning that these targets have connections with more unique word nodes (M_{High} = 203.3 vs. M_{Low} = 188.8; $t(11) = 6.05$, $p = < .001$). Having connections with more unique word nodes allows more activations to circulate back to the target node, thereby reinforcing its activation (Mak et al., 2020; Vitevitch, Chan, & Roodenrys, 2012). Nevertheless, this was also the case in Simulation 1, where the number of unique word nodes was also higher in high-diversity networks (M_{High} = 276.6 vs. M_{Low} = 237.1; $t(7) = 5.28$, $p = .001$). Given Simulation 1 did not find a high-diversity advantage, it suggests that being connected with more unique word nodes alone cannot account for the high-diversity advantage in Simulation 2b. One explanation is that in addition to being connected to more unique word nodes, the target words in the high-diversity networks in Simulation 2b also had a relatively close-knit immediate neighborhood, established as a result of the anchoring passages. This appears to be the case given the mean clustering coefficients of the high-diversity words increased from 0.206 in Simulation 1 to 0.231 in Simulation 2b. Overall, it is reasonable to attribute the high-diversity advantage found in Simulation 2b to two factors: (i) an anchoring phase that produced more close-knit neighborhoods compared to those in Simulation 1, and (ii) an opportunity for building associations with more unique word nodes, which further helps activations to circulate back to the targets.

Finally, it is worth comparing the low-diversity conditions in Simulations 1 and 2b. In Simulation 1, the target words in the low-diversity condition were anchored onto one topic six times while in Simulation 2b, they were anchored onto one topic eight times. If anchoring was always beneficial to word learning, one might expect the words in the low-diversity networks of Simulation 2b to receive more activations than those in the high-diversity networks. However, the opposite was true in Simulation 2b. In light of this, I suggest that the effect of anchoring may *saturate* (e.g., Murray & Forster, 2004); that is, repeated anchoring beyond a certain threshold might have little or even no detectable effect on subsequent processing. In turn, this suggests that while anchoring (i.e., experiencing new words in similar and familiar contexts) may be conducive to the initial stage of word learning, additional anchoring beyond a certain threshold may lead to little changes in the learning outcomes.

Finally, caution is required when interpreting the results of Simulation 2b: This kind of network simulation does not take into account the effect of learning order, meaning that Simulation 2b would have produced the same result had the order of topics in the high-diversity condition been intermixed (i.e., Topic 11111222 and Topic 12211211 would

produce identical simulation results). Therefore, while Simulation 2b provides evidence for anchoring being a potential contributor to a high-diversity advantage in word learning, it does not speak to whether anchoring must precede diversity. In line with our account, however, a computational study by Jorge-Botana, Olmos, and Sanjosé (2017) suggested that anchoring preceding diversity might be the most conducive growth path for word knowledge. Future experimental work is needed to determine the importance of learning order.

To sum up, Simulation 2 sets up the prediction that participants should show a recognition advantage for words experienced in diverse contexts, if the words were given an opportunity to anchor onto one discourse topic. This prediction was tested in Experiment 2.

Experiment 2

Anchoring a word before introducing diversity

2. Introduction

As in Experiment 1, young adults read passages containing novel words in high vs. low semantic diversity contexts. The reading materials were taken from Simulation 2. Unlike Experiment 1, the target words in the high-diversity conditions were given an anchoring opportunity (i.e., repeatedly experiencing a word in a familiar topic) in the first five passages before diversity was introduced, following the procedure in Simulation 2 (see Table 3.1). In light of the results from Simulation 2, my hypothesis was that, given an anchoring opportunity, target words experienced in the high-diversity condition would show a processing advantage in visual recognition. Specifically, it was predicted that they would be recognized more efficiently in speeded recognition (greater accuracy and/or faster RTs) and that they would induce greater top-down interference in the pseudoword inferiority task, leading to poorer performance. As for semantic judgement, I predicted that high-diversity novel words would be judged more poorly than the low-diversity words, in line with Experiment 1 and Johns et al. (2016). Experiment 2 was pre-registered ahead of data collection (https://aspredicted.org/blind.php?x=ti8g6f).[5]

[5] The pre-registration document included an additional post-test, namely importance judgement. As this did not feature in Experiment 1, I chose not to report it here to avoid over-complication.

2.1. Method

2.1.1. Participants

Forty-three undergraduate psychology students from Oxford University (11 males; M_{age} = 19.4; SD_{age} = 1.49) participated for course credits. All were native speakers of English with normal or corrected-to-normal vision and all reported no known history of any developmental disorders. None of them had taken part in Experiment 1. Following the pre-registered exclusionary criteria, three participants were excluded: two because they scored 2 SDs below the sample mean on the comprehension questions, and one because their accuracy rates in the speeded recognition task fell below 75%. All analyses, unless otherwise specified, were based on the remaining 40 participants.

2.1.2. Materials and procedure

Experiment 2 used the reading materials developed for Simulation 2. As in Experiment 1, the 12 target base words (e.g., *mendacious*) were replaced by pronounceable pseudowords (e.g., *mendacious* → ossanic). The general procedure also followed Experiment 1. Participants first completed an exposure phase where they read the passages; they then completed a surprise test phase designed to measure target word learning.

As per Experiment 1, semantic diversity was manipulated within-subject. Each participant saw a total of 12 novel words, six in the high-diversity condition and six in the low-diversity condition. Assignment to the diversity conditions was fully counterbalanced across items. Participants saw each item eight times, five times in the anchoring phase and three times in the post-anchoring phase. The diversity manipulation was induced in the post-anchoring phase; prior to this, the experience of high vs. low diversity words was identical in the anchoring phase. Note that the high-diversity conditions in Experiment 1 and 2, despite sharing the same name, were somewhat different from each other (i.e., six topics in Experiment 1, but only two topics in Experiment 2).

(i) Exposure phase

The exposure phase in Experiment 2 had two components: (i) anchoring passages and (ii) post-anchoring passages. Each participant read a total of 60 anchoring passages and then 18 post-anchoring passages. For each anchoring passage, participants answered one yes/no question. Pilot testing confirmed that question difficulty was equivalent across the diversity conditions. Each passage contained one exposure to one target word. For the post-anchoring passages, participants rated the comprehensibility of each passage and answered

two comprehension questions (taken from Experiment 1). Each of these post-anchoring passages contained two target words, one belonging to the high-diversity condition, another to the low-diversity condition. All participants read the same 18 post-anchoring passages.

(ii) Test phase

Learning was assessed with three post-tests, namely pseudoword inferiority, speeded recognition, and semantic relatedness judgement. Task parameters and procedures were identical to Experiment 1.

2.2. Results

The analysis approach followed the pre-registered analysis plan and was identical to that described in Experiment 1.

(i) Exposure phase

On average, participants achieved 95.0% (*SD* = 2.13%) and 82.2% (*SD* = 2.93%)[6] accuracy in the comprehension questions associated with the anchoring and post-anchoring passages, respectively. There was no accuracy difference between the high- and low-diversity passages in the anchoring phase, $t(39) = 0.28$, $p = .89$. In the post-anchoring phase, since each passage contained one high-diversity and one low-diversity novel word, it was not possible to compare the accuracy rates across the two diversity conditions.

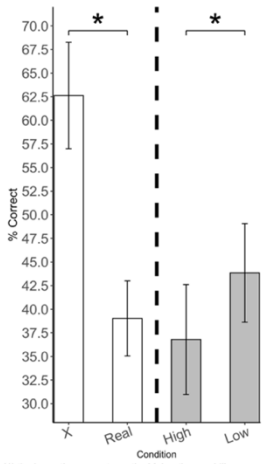
(ii) Test phase

Pseudoword inferiority. Mean accuracy in the four conditions is summarized in Figure 3.4. Letter identification in the real word condition (*M* = 37.9%) was significantly poorer than in the X condition (*M* = 60.7%) ($z = 3.26$, $p = .001$), replicating the pseudoword inferiority effect reported by Grainger and Jacobs (2005) and observed in Experiment 1. Letter identification was also significantly poorer in both diversity conditions than the X condition (zs > -4, ps < .001). This indicates that the novel words in both conditions had been learned sufficiently to bias letter identification. Finally, there was also a significant difference between these two diversity conditions, with novel words in the high-diversity passages (*M* = 35.7%) biasing letter identification to a greater extent than those in the low-diversity passages (*M* = 42.5%; $z = -2.42$, $p = .016$; see Table 3.2 for model summary).

[6] This difference is likely a consequence of the anchoring passages (*M number of words* = 36 words) being shorter in length (and hence less demanding) than the post-anchoring passages (*M number of words* = 127 words).

Figure 3.4

Mean accuracy rates across the four conditions in the pseudoword inferiority task, Experiment 2.



Note. (i) Error bars represent 95% within-subject confidence intervals. (ii) Lower accuracy in the real and high-diversity conditions is indicative of greater top-down interference.

Table 3.2

*Summary of the GLME models examining the effect of semantic diversity on **accuracy** in each of the post-tests, Experiment 2.*

	Estimate	Standard Error	z value	p value
(a) Pseudoword inferiority				
Intercept	-0.52	0.21	-2.44	.015*
SemD	-0.17	0.07	-2.42	.016*
(b) Speeded recognition				
Intercept	3.51	0.30	11.64	<.001*
SemD	0.08	0.25	0.30	.762
(c) Semantic relatedness				
Intercept	1.24	0.16	7.70	<.001*
SemD	0.08	0.06	1.43	.152

Table 3.3

*Summary of the LME models examining the effect of semantic diversity on **RTs** in each of the post-tests, Experiment 2.*

	Estimate	Standard Error	t value	p value
(a) Speeded recognition				
Intercept	-153.2	3.04	-50.34	<.001*
SemD	-2.21	0.75	-2.92	.004*

(b) Semantic relatedness

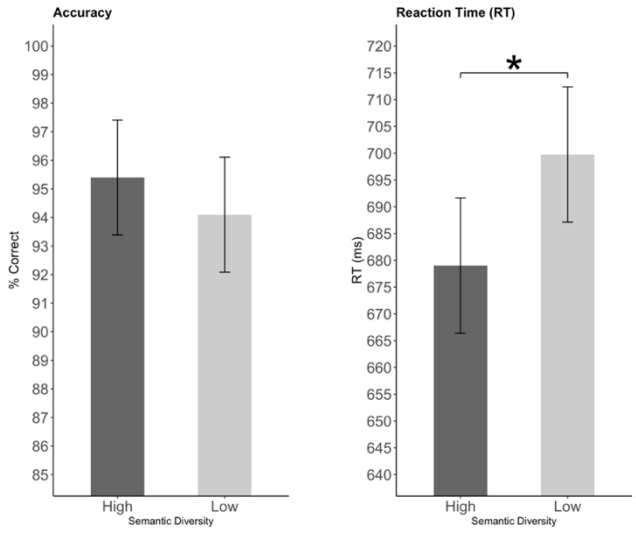
Intercept	718.68	2.34	307.21	<.001*
SemD	-0.92	0.97	-0.94	.36

Note. To improve interpretability, estimate and standard error were multiplied by 100,000 in speeded recognition and by 100 in semantic relatedness.

Speeded recognition. Data from two participants were removed as they appeared to use the wrong response keys; this analysis is therefore based on 38 participants. Mean accuracy rates in the two diversity conditions are summarized in the left panel of Figure 3.5. Target words in the high-diversity condition (*M* = 95.4%) were recognized more accurately than those in the low-diversity condition (*M* = 94.1%), but this was not statistically significant (*z* = -0.30, *p* = .762; see Table 3.2 for model summary). Turning to RTs, mean RTs (to correct 'yes' trials) in the two diversity conditions are summarized in the right panel of Figure 3.5. Participants were significantly faster at recognizing words that were encountered in the high-diversity passages (*M* = 679 ms) than those in the low-diversity passages (*M* = 700 ms; *t* = -2.92, *p* = .004, see Table 3.3 for model summary).

Figure 3.5

Mean accuracy (left) and RTs (right) in the two semantic diversity conditions in speeded recognition, Experiment 2.

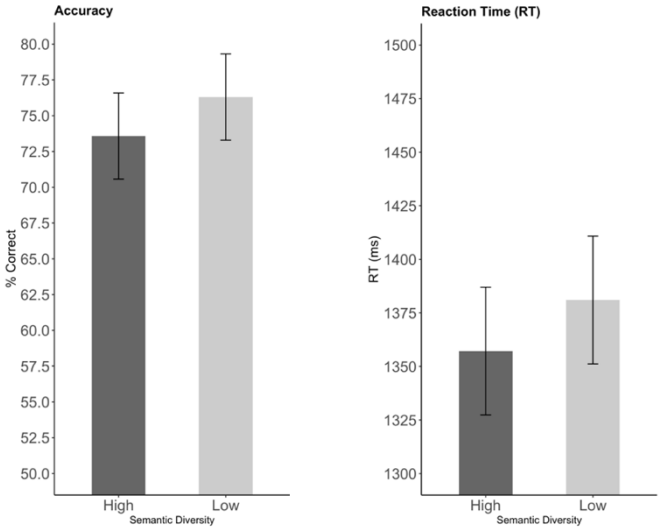


Note. Error bars represent 95% within-subject confidence intervals.

Semantic relatedness. Overall accuracy was good (*M* = 75%, *SD* = 14.7%) and significantly above chance, $t(39) = 9.63$, $p < .001$. Mean accuracy rates in the two diversity conditions are plotted in the left panel of Figure 3.6. The corresponding GLME model, summarized in Table 3.2, found no reliable difference between the two diversity conditions ($z = 1.43$, $p = .152$). Mean RTs (to correct trials) in the two diversity conditions are plotted in the right panel of Figure 3.6. While target words in the high-diversity condition were judged more quickly, the difference was not significant ($t = -0.94$, $p = .36$; see Table 3.3 for model summary).

Figure 3.6

Mean accuracy (left) and RTs (right) in the two semantic diversity conditions in semantic relatedness judgement, Experiment 2.



Note. Error bars represent 95% within-subject confidence intervals.

2.3. Discussion

Experiment 2 examined the effect of semantic diversity on incidental word learning during independent reading. Young adults read passages containing pseudowords (e.g., OSSANIC) whose meanings were derived from words such as *mendacious.* Experiment 2 differed from Experiment 1 in one major respect: In Experiment 2, pseudowords in the high-diversity condition were experienced in one familiar topic in the anchoring passages before diversity was introduced. Items experienced in this condition (vs. those in the low-diversity condition) showed a greater pseudoword inferiority effect and were recognized faster in speeded recognition. Although the mean RT advantage enjoyed by the high-diversity words in speeded recognition was small at 21 ms, it is comparable with the 26-ms advantage

reported by Johns et al. (2016). These findings also align with the data reported in Simulation 2, which showed that diversity, coupled with an anchoring opportunity, could boost the amount of activation received by the target words.

Why did the results here contrast with those from Experiment 1?

The findings from Experiment 2 contrast with those from Experiment 1, which demonstrated a clear low-diversity advantage in both the learning of forms and meanings. First, consider the low-diversity condition in Experiment 2: I proposed that the benefits associated with anchoring on subsequent processing may saturate (Jorge-Botana et al., 2017; Murray & Forster, 2004), meaning that repeated anchoring beyond a certain threshold may result in little or even no detectable effect on behavioral measures. This is particularly evident in the accuracy data from speeded recognition across Experiments 1 and 2. In Experiment 1, novel words in the low-diversity condition were anchored onto one topic six times, while in Experiment 2, they were anchored onto one topic eight times. Despite the extra anchoring (and hence extra exposure), mean accuracy in speeded recognition was almost identical (94.4% in Exp 1 vs. 94.1% in Exp 2).[7] This provides credence for the notion that the effect of anchoring on word learning may saturate, meaning that a low-diversity advantage may be more apparent in the earliest stage of word learning but may plateau or even decline as learning progresses. Of course, saturation may also apply to contextual variation, but its trajectory may be different. Future studies should use computational methods to compare the saturation trajectory for anchoring and diversity (see e.g., Jorge-Botana et al., 2017).

Next, turning to the high-diversity condition: One of the critical differences between the high-diversity conditions in Experiment 1 and 2 is that in the latter, the high-diversity words were initially given an anchoring opportunity (i.e., being repeatedly experienced in one familiar topic such that it could be secured onto long-term memory). This might have helped stabilize and strengthen the mental representations of the novel words (Jorge-Botana et al., 2017; Mak, 2019), allowing them to be better placed to take advantage of contextual variation subsequently. However, given results from Johns et al. (2016) and Rosa et al., (2017), anchoring should *not* be interpreted as a pre-condition for the high-diversity advantage seen in Experiment 2; instead, it should be conceived of as a facilitator or a contributor, which enabled the novel words to be better placed to take advantage of contextual variation subsequently.

[7] In the high-diversity conditions, mean accuracy increased from 89.3% (Exp 1) to 95.4% (Exp 2).

While an anchoring account fits with the high-diversity advantage here, there are at least two other plausible explanations. First, high-diversity words in Experiment 1 were experienced in six topics while in Experiment 2, they were only experienced in two topics. It is plausible that a little bit of diversity, as in Experiment 2, is conducive to word learning, but too much diversity, as in Experiment 1, is not. While this possibility cannot be ruled out, it seems to be at odds with corpus-based studies (e.g., Hsiao & Nation, 2018) and the results from Johns et al. (2016), where the high-diversity novel words were advantaged in visual recognition, despite being experienced in five distinct topics. Second, high-diversity words in Experiment 1 were experienced once in each topic, but in Experiment 2, they were experienced five or three times in each topic, so perhaps high semantic diversity is only beneficial when readers have multiple opportunities to map the word to each topic. This is plausible, but again, it appears at odds with the high-diversity advantage reported in Johns et al. (2016), where the high-diversity words were only experienced once in each topic. Admittedly, the design discrepancy between Experiments 1 and 2 does not allow the ruling out of these alternative explanations, so future work is necessary before it could be concluded with certainty that it was anchoring preceding diversity that drove the high-diversity advantage observed in Experiment 2. However, it is difficult to see how the alternative accounts could provide a neat explanation to the findings here as well as to those from Johns et al. (2016). Therefore, at present, I contend that the anchoring account provides a more convincing explanation to the result discrepancy between Experiments 1 and 2.

How does contextual variation boost the learning of word form?

As to why the high-diversity words induced greater pseudoword inferiority effect and were identified more efficiently in Experiment 2, there are multiple related explanations. The first one stems from the Semantic Distinctiveness Model (Jones et al., 2012; Jones et al., 2017; see also Nelson & Shiffrin, 2013), which postulates that when a word is experienced, the encounter will be compared to its contextual history stored in memory; if the current context is distinct from the word's contextual history, there is more pressure for stronger encoding (Bolger et al., 2008; Johns et al., 2016) and perhaps more pressure for active retrieval of previous encounters. A by-product of this may be more well-specified and consolidated mental representations (e.g., Roediger & Butler 2011).

Second, a word that has been experienced in multiple contexts may become more context-independent (i.e., being less associated with a specific contexts). As others have

argued, a behavioral marker of greater context-independence might be enhanced word identification in isolation (Hsiao et al., 2020) and in new contexts (Pagán et al., 2020). What follows is that the learning opportunity afforded by an encounter in a diverse context is more valuable than one in a repeating or highly similar context (e.g., Jones et al., 2012; Kachergis, Shiffrin, & Chen, 2009; Lev-Ari, 2016; Nation, 2017).

Third, and relatedly, the high-diversity words in Experiment 2 might have become more polysemous as they were experienced in an additional topic. Polysemous words in general are recognized faster than non-polysemous words in lexical decision (Rodd et al., 2002; 2004), plausibly due to the former receiving more supportive activation from their distinct but overlapping senses (see also Floyd & Goldberg, 2020). This explanation fits with the view that semantic diversity is a continuous measure of word senses (Hoffman & Woollams, 2015). However, Cevoli et al. (2020) recently challenged this view by showing that semantic diversity values derived from corpus analysis do not capture variation in meaning, but variation in topics. Future work is therefore needed to delineate what it means for a word to be polysemous and how this relates to contextual experience.

No evidence that semantic diversity influenced the learning of meaning in Experiment 2

There was no effect of semantic diversity on the learning of meaning in Experiment 2, as shown by both the accuracy and RT data from semantic relatedness. This contrasts with the findings of Experiment 1 where low-diversity words were easier to judge (see also Johns et al., 2016). This discrepancy may be explained by two accounts. First, in Experiment 2, words in the high-diversity condition were experienced in two topics only, compared with six topics in Experiment 1 (and five topics in Johns et al., 2016). Potentially, therefore, the meanings of the high-diversity words in Experiment 2 had not become complex or nuanced enough to hinder performance. Second, the high-diversity words in Experiment 2 were experienced in each of the two topics five or three times while those in Experiment 1 were experienced in each of the six topics once. In other words, readers in Experiment 2 had more opportunities to map the high-diversity novel words onto each discourse topic. Potentially, the high-diversity words in Experiment 2 may become better integrated into the two topics (i.e., pre-existing semantic knowledge), allowing them to be judged as well as the low-diversity words. This possibility is evaluated mathematically in the next chapter.

Of note is that while relatedness judgement is a widely adopted paradigm in psycholinguistics, interpreting the results can be tricky in the context of word learning. If for example high-diversity novel words were judged more poorly (as in Experiment 1), two

interpretations are possible: (i) the meanings of the novel words were poorly learned (hence poor task performance), and (ii) the meanings of the novel words were learned fairly well but their complex contextual history made semantic judgement difficult. As such, future studies interested in how word learning is influenced by contextual variation should consider using a range of tasks to tap semantic knowledge to complement relatedness judgement (e.g., meaning generation; Bolger et al., 2008; sentence plausibility; Experiment 4 in Chapter 6; Joseph & Nation, 2018).

2.3.1. Limitations

The focus of Chapters 1 and 2 has been with investigating the consequences of manipulating semantic diversity on word learning while other important factors such as frequency were held constant. Any difference between the high and low diversity conditions within each experiment cannot be a simple consequence of differences in frequency (or document count) as it was equated across conditions. However, there was a frequency difference *between* experiments. In Experiment 1, each novel word was seen six times (once in six different passages) whereas in Experiment 2, each word was seen once in eight different passages. The amount of variation also differed across the two experiments. In Experiment 1, high-diversity contexts were sampled from six different discourse topics whereas in Experiment 2, variation came from switching the discourse topic once. This was an active decision to keep the experiment within a single session and therefore not complicated by multi-session learning. Ultimately, of course, word learning is typically an incremental and dynamic process spanning a prolonged period of time. Future studies should consider multi-day training paradigms and the effect of number of discourse topics, not only to tackle more words and more varieties of variation, but also to consider how these factors interact with other factors known to influence learning and memory (e.g., spacing and testing effects; Hulme, 2018) as well as sleep-related consolidation (e.g., James, Gaskell, & Henderson, 2019; Henderson, Weighall, & Gaskell, 2013). Relatedly, learning was measured immediately after exposure; there is a clear need to investigate longer-term retention via delayed post-tests (Experiment 4, Chapter 6). Finally, while both Simulation 2 and Experiment 2 suggested that anchoring helped bring about the emergence of a high-diversity advantage in the learning of form, the experimental design does not allow me to conclude with certainty that anchoring must precede diversity. Further experimental work is required to determine the importance of learning order, although existing findings on learning words (Jorge-Botana et al. 2017), and aspects of morphology (Tamminen, Davis, & Rastle, 2015), and syntax (Elman,

1993) all converge to suggest that starting small (i.e., restricting diversity) may facilitate subsequent learning.

Chapter 4

Simulation 3

Using word vectors to simulate the learning of meanings across Experiments 1 and 2

1. Introduction

Simulations 1 and 2 examined the effect of semantic diversity on the learning of *form* by simulating spreading activation in lexical networks. In this chapter, I present Simulation 3 where I examined the effect of semantic diversity on the learning of *meaning* using word vectors.

In Experiment 1 (Chapter 2), novel words experienced in the high-diversity condition elicited less accurate and slower performance in a semantic relatedness judgement task than those experienced in the low-diversity condition. In Experiment 2, however, this disadvantage disappeared and words experienced in the two diversity conditions were judged with comparable efficiency. Two plausible, albeit tightly related, explanations were proposed. First, the high-diversity words in Experiment 2 were experienced in two topics only, compared with six topics in Experiment 1. The meanings of the high-diversity words in Experiment 2, therefore, may not have grown nuanced enough to hinder performance on relatedness judgement. In this chapter, I referred to this account as the *semantic complexity account*, which has been the "go-to" explanation to why high-diversity words are often disadvantaged in semantic judgement (e.g., Chapman, 2019; Johns et al., 2016; Jones et al., 2012; Hoffman & Woollams, 2015).

A second account is also possible. The high-diversity words in Experiment 1 were experienced in each of the six topics once while in Experiment 2, they were experienced in each of the two topics five or three times. In other words, readers in Experiment 2 had more opportunities to map the high-diversity novel words onto each familiar discourse topic. Therefore, the high-diversity words in Experiment 2 may become better integrated into pre-existing long-term knowledge than those in Experiment 1, allowing them to be judged as well as the low-diversity words. This is referred to the *integration account*. In this chapter, I evaluated this account mathematically by calculating word vectors from the reading materials I developed for Experiments 1 and 2. With these word vectors, I am able to examine how semantic diversity influences a word's semantic space, and by inference, the learning of meanings.

1.1. What is vector semantics?

In a nutshell, vector semantics uses a row of numbers to represent a word's meaning, computed based on the word's distribution in texts. According to Jurafsky and Martin (2019), its origin can be traced back to 1950s, when Osgood, Suci, and Tannenbaum (1957) found that words vary along three dimensions of affective meaning: *valence*, *arousal*, and *dominance*, and that a word's meaning can be represented by subjective ratings on each of these three dimensions, illustrated in Table 4.1. Osgood et al. (1957) reported that using these three numbers, a word's meaning can be represented by a point in a three-dimensional space, where each dimension corresponds to the word's values on these affective scales. This ground-breaking idea formed the foundation of many techniques in natural language processing, including vector semantics.

Table 4.1

Subjective ratings of valence, arousal, and dominance of six words. Data were taken from the English Lexicon Project (Balota et al., 2007).

Word	Valence	Arousal	Dominance
desk	5.56	2.45	5.72
table	5.49	3.00	5.84
great	7.50	4.14	6.65
wonderful	7.41	4.57	7.19
sad	2.10	3.49	3.84
sorrow	2.95	3.55	3.80

Around the same time as Osgood et al. (1957), linguists such as Firth (1957) and Wittgenstein (1953) proposed that words that co-occur with similar words and contexts should have related meanings. This insight led to the famous statement, "You shall know a word by the company it keeps" (Firth, 1957). To illustrate this distributionalist approach to semantics, consider the Cantonese word *gaichoi* in the following sentences:

1. Gaichoi should be cooked with ginger.
2. Gaichoi is very nutritional.
3. Vegetarians should try gaichoi.

And let's suppose you had encountered these sentences before:

4. Vegetarians love spinach.
5. Spinach packs high amounts of vitamin C and K, and therefore is very nutritional.
6. Spinach should be cooked in boiling water.

As *gaichoi* and *spinach* both co-occur with words like *nutritional*, *vegetarian*, and *cook*, this leads to the inference that *gaichoi* is also an edible leafy vegetable. This distributional approach to word meaning can be implemented computationally by tracking co-occurrence in natural language and using this to quantify degree of semantic relatedness between any pairs of words (Jurafsky & Martin, 2019).

Fast forward to the early 21st century, advances in computing capability helped marry the works of Osgood et al. (1957) and of distributional linguists like Firth (1957), giving rise to various methods to compute word vectors, including two popular machine learning algorithms known as word2vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014). Although each method is unique in how they define the numbers in the word vector, they are based on counts of surrounding words such that words co-occurring with similar words will be similar in terms of the values in their word vectors and will therefore be mapped in close proximity to each other in an N-dimensional semantic space. For instance, words such as *orange* should be close to *apple* in a semantic space but relatively far away from *lamppost.* A striking aspect of word vectors is that they lend themselves to mathematical operators, such that if inputted [queen – woman + man], one would obtain the vector for [king] (Mikolov et al., 2013).

In Simulation 3, I ask whether it is possible to use word vectors to simulate the learning of word meanings, using the materials from Experiments 1 and 2 as language input. To this end, words in the reading materials from the two experiments were transformed into word vectors, and then mapped onto a semantic space. This allowed the quantification of the degree of semantic relatedness between a word target and the words with which it co-occurs. Recall that the word targets were unknown words (i.e., participants learned *ossanic* which represented a concept aligned with *mendacious*) in the context of highly familiar neighboring words (e.g., *Trump*, *tweet*). I inferred that the closer the word targets are to the neighboring words in semantic space, the greater the likelihood that targets would be integrated into pre-existing semantic knowledge. In turn, this should facilitate the ease with which their meanings are learned and judged in semantic relatedness.

2. Simulation 3

2.1. Overview

Two sets of simulations were conducted: Simulation 3a was based on the reading materials from Experiment 1 while Simulation 3b on the reading materials from Experiment 2. There are many ways to obtain word vectors, including but not limited to machine learning

techniques such as CBOW and Skip-gram in word2vec (Mikolov et al., 2013). However, previous studies (e.g., Altszyler, Sigman, Ribeiro, & Slezak, 2016) have demonstrated that these prediction-based techniques require very large corpora (e.g., over 1 million words) to produce good-quality vectors. More appropriate, given that nature of my reading materials, is the count-based approach developed by Silge (2017). This approach makes use of (co-)occurrence statistics and matrix factorization to produce word vectors and is more well suited for smaller language corpora.

2.2. Method

The 16 corpora (8 novel words x 2 diversity conditions) prepared for Simulation 2a were used in Simulation 3a while the 24 corpora (12 novel words x 2 diversity conditions) prepared for Simulation 2b were used in Simulation 3b. Both simulations were conducted in R (R Core Team, 2019), following the same procedure outlined by Silge (2017). Two R packages were used: tidytext (Silge & Robinson, 2016) and irlba (Baglama, Reichel, & Lewis, 2019). All code is available on the Open Science Frameworks (https://osf.io/a57bn/?view_only=160f92cf350c472e8a968bfa0106a6ae).

After lemmatization and removing function words, I calculated how many times a word appeared in a corpus and its unigram probability (i.e., number of occurrences divided by the total number of words in a corpus). Next, I calculated a word's skipgram probability (i.e., how often a word co-occurs with another word). This procedure uses a fixed-size sliding window that centers around each word. For instance, applying a window size of four to a sentence like "Psychology ~~is the~~ scientific study ~~of the~~ mind, behavior ~~and~~ brain" means that after removing the function words, the word *mind* has *scientific*, *study*, *behavior*, and *brain* as co-occurring neighbors; if the window size is reduced to two, it will only have *study* and *behavior* as its neighbors. In this simulation, I used four different window sizes (i.e., 8, 12, 16, 20) to check whether the results are generalizable. I set the minimum window size at eight as my corpora are small (~500 words per corpus); using small windows would severely limit the variety of neighbors a target word could have.

The skipgram probabilities were then normalized to see the extent to which two words co-occur more often than expected, given their unigram frequency. After normalization, the skipgram probabilities were cast to a sparse matrix to enable matrix factorization. Then, the matrix was reduced in dimensionality using Singular Value Decomposition (Deerwester et al., 1990). Two values needed to be determined: number of singular vectors and maximum number of iterations. There is no general consensus as to

what the ideal values are (Silge, 2017), so I used the default values (5 and 1000 respectively) in the irlba package. After this, two sets of semantic spaces were created for each target word (one for the high-diversity condition, the other for the low-diversity condition). This enabled the degree of semantic relatedness between the target and its neighboring associates in each diversity condition to be quantified. Overall, Simulation 3a produced 64 semantic spaces (8 target words x 2 diversity conditions x 4 window sizes) while Simulation 3b produced 96 (12 target words x 2 diversity conditions x 4 window sizes).

2.3. Results

As a reminder before going further: I used the base words (e.g., *mendacious*) as the word targets in the simulations here, although participants in Experiments 1 and 2 were exposed to pseudowords (e.g., *ossanic*) which replaced the base words in the reading passages.

The associates of each target word were ranked in descending order according to their degree of relatedness to the target word. The larger the relatedness value, the closer the associate is to the target in semantic space. As an example, Table 4.2 summarizes the top 10 associates to the base word *mendacious* when it appeared in the high- and low-diversity conditions in Simulation 3a (window size = 16).

Table 4.2

Ten nearest associates to the word "mendacious" when it appears in the high- and low-diversity conditions in Simulation 3a (window size = 16).

Word target: Mendacious			
High Semantic Diversity		Low Semantic Diversity	
Associates	Relatedness Value	Associates	Relatedness Value
Change	0.0907	*Person*	0.444
Colour	0.0826	*Trump*	0.0938
Express	0.0543	*Popular*	0.0283
Response	1.29e-7	*Power*	0.0246
Meet	5.46e-8	*Call*	0.0162
Time	2.97e-8	*Demagogue*	0.0111
Effort	2.98e-13	*Observe*	0.0110
Russian	2.57e-13	*Mainstream*	0.0105
Good	2.09e-13	*Want*	0.00602
Announce	1.60e-13	*Balkin* (a last name)	0.00221
Neighborhood relatedness score	0.0228	Neighborhood relatedness score	0.0647

Next, I averaged the relatedness values of the top 10 associates to obtain a neighborhood relatedness score for each target word.[8] From this, I inferred that the greater the relatedness score, the more likely the associated neighborhood (composed mainly of familiar concepts for human learners) can support the integration of the new word's meaning into the familiar topic. In turn, this should correspond to better performance in the semantic relatedness judgement task.

At each of the four window sizes I used, each target word had two neighborhood relatedness scores: one for the high-diversity condition, the other for the low-diversity condition. These two scores were then compared to see how semantic diversity might influence the learning of meanings, and by inference, human performance on the relatedness judgement task reported in Experiments 1 and 2.

2.3.1. Simulation 3a: reading material from Experiment 1

Table 4.3 summarizes the neighborhood relatedness scores from the 64 semantic spaces derived from Simulation 3a, based on the reading materials used in Experiment 1. At each window size, the mean relatedness score was higher in the low-diversity condition than the high-diversity condition. This pattern was seen for 26 (or 81.25%) of the 32 pairs.

A Wilcoxon signed-rank test was conducted at each window size (with Bonferroni correction), showing that the relatedness values were significantly higher in the low-diversity condition than in the high-diversity condition (ps < .001). These findings are consistent with the notion that target words in the low-diversity condition become more associated with pre-existing concepts, and as a consequence, they are more likely to be integrated into long-term memory than those in the high-diversity condition.

[8] Choosing 10 associates because it included a variety of words but at the same time excluded words with extremely small relatedness values (e.g., associates beyond the 11th rank tended to have extremely small relatedness values).

Table 4.3

A summary of the neighborhood relatedness scores from the 64 semantic spaces derived from Simulation 3a.

Window size	8		12		16		20	
Target words	High SemD	Low SemD	High SemD	Low SemD	High SemD	Low SemD	High SemD	Low SemD
Latent	0.00013	0.01762	7.59e-09	7.69e-05	0.00101	0.02673	0.01248	0.02119
Constellation	1.11e-16	1.76e-14	5.23e-13	0.07800	0.00496	0.09810	0.06445	0.10870
Predilection	2.29e-12	8.525e-2	0.04947	0.04423	0.03399	0.02929	0.04969	0.01343
Mendacious	6.13e-11	1.45e-09	1.29e-09	0.05207	0.02276	0.06476	0.05925	0.09652
Diffident	1.04e-12	1.20e-10	0.04992	0.01747	6.60e-10	0.11839	0.01520	0.13759
Proliferation	1.10e-10	0.05172	2.16e-10	0.07756	0.01578	0.09849	0.00974	0.10064
Vagarious	2.11e-10	0.05167	0.06478	0.06068	0.02245	0.09548	0.01121	0.10882
Castigate	1.66e-11	1.89e-10	0.00482	0.00920	0.00712	0.00415	0.00756	0.01435
Mean	1.63e-05	0.02578	0.02112	0.04241	0.01351	0.06692	0.02870	0.07516
Wilcox sign-ranked test	$Z(79) = -6.35$, $p < .001$*		$Z(79) = -5.88$, $p < .001$*		$Z(79) = -6.30$, $p < .001$*		$Z(79) = -5.80$, $p < .001$*	

Note. (i) * denotes statistically significant results. (ii) Cells in gray are pairs where the low-diversity condition had a higher neighborhood relatedness score.

2.3.2. Simulation 3b: reading materials from Experiment 2

Table 4.4 summarizes the neighborhood relatedness scores from the 96 semantic spaces derived from Simulation 3b. A comparison between diversity conditions within each window size showed that the relatedness scores are higher in the low-diversity condition in 21 (or 43.75%) of the 48 pairs.

Following the previous analysis, a Wilcoxon signed-rank test (with Bonferroni correction) was conducted at each window size. At a window size of 8, the relatedness values were significantly higher in the low-diversity condition [$Z(119) = -4.83$, $p < .001$]. In contrast, when the window size increased to 16, the direction of effect flipped such that the relatedness values were higher in the high-diversity condition [$Z(119) = 2.93$, $p < .001$]. To complicate the picture further, there was no significant difference between the diversity conditions when the window size was 12 or 20 (ps > .0125). Taken together, Simulation 3b provided no consistent result as to whether the high- and low-diversity target words in Experiment 2 were different in their degree of relatedness to their surrounding words. By inference, target words in both conditions might be similar in terms of their likelihood to be integrated into pre-existing semantic knowledge by human learners. This interpretation fits with the finding that high- and low-diversity novel words were judged with comparable efficiency in relatedness judgement in Experiment 2.

Table 4.4

A summary of the neighborhood relatedness scores from the 96 semantic spaces derived from Simulation 3b.

Window size	8		12		16		20	
Target words	High SemD	Low SemD	High SemD	Low SemD	High SemD	Low SemD	High SemD	Low SemD
Latent	0.00599	0.00487	0.00772	0.00677	0.01148	0.01062	0.01081	0.00963
Constellation	0.02910	0.05552	0.02183	0.01507	0.01826	0.01566	0.01473	0.01534
Predilection	0.01267	0.02009	0.01294	0.01710	0.01091	0.01544	0.01152	0.01494
Mendacious	0.01105	0.04162	0.02155	0.02102	0.01755	0.01803	0.01690	0.01749
Diffident	0.02304	0.02086	0.02133	0.02096	0.01774	0.01766	0.016944	0.01694
Proliferation	0.01269	0.01915	0.01617	0.01516	0.01441	0.01311	0.01408	0.01335
Vagarious	0.04280	0.04433	0.01439	0.01396	0.01175	0.01186	0.01086	0.01193
Castigate	0.01194	0.00696	0.01237	0.01301	0.00967	0.01102	0.01244	0.01318
Imbroglio	0.00845	0.00431	0.00709	0.00162	0.00284	0.00023	0.00207	1.08e-06
Architecture	0.00812	0.00709	0.00908	0.01096	0.01281	0.01081	0.01064	0.00951
Shenanigans	0.01742	0.02468	0.01590	0.01604	0.00915	0.00855	0.01315	0.01093
Avidity	0.00519	0.02140	0.01738	0.01805	0.01747	0.01692	0.01808	0.01706
Mean	0.01571	0.02257	0.01481	0.01414	0.01284	0.01250	0.01269	0.01253
Wilcox sign-ranked test	$Z(119) = -4.83$, $p < .001$*		$Z(119) = -1.77$, $p = .038$		$Z(119) = -2.93$, $p = .001$*		$Z(119) = -2.07$, $p = .019$	

Note. (i) * denotes statistically significant results. (ii) Cells in gray are pairs where the low-diversity condition had a higher neighborhood relatedness score.

2.4. Discussion

Simulation 3 investigated semantic diversity influences on the learning of meaning, using the reading materials developed for Experiments 1 and 2. Specifically, I tested whether semantic diversity influences the degree to which a word target is associated with its surrounding words. Word vectors were derived from the two sets of reading materials, and from these I quantified the degree of relatedness between a target word and its surrounding words. I inferred that a target word might be better integrated into the discourse topic when it is closer to its surrounding words in semantic space.

Simulation 3a, built upon the reading materials from Experiment 1, found that target words in the low-diversity condition were consistently closer to the surrounding words in semantic space than those in the high-diversity condition. This suggests that a low-diversity word might be more able to integrate into pre-existing knowledge than a high-diversity word. This computational finding aligns with the behavioral data from Experiment 1 where novel words experienced in the low-diversity condition were easier to process in semantic relatedness. In contrast to the consistent pattern across items and across window sizes in Simulation 3a, data from Simulation 3b were less consistent. There was no overall difference in relatedness scores across the two diversity environments, mirroring the lack of difference in performance on the semantic judgement task across the two conditions for human participants in Experiment 2. These simulations provide computational evidence for the possibility that semantic diversity modulates the learning of meaning by affecting the integration of a target word into the discourse topic.

It is generally accepted that words appearing in more diverse contexts tend to have more nuanced semantic representations (Jones et al., 2012; Hoffman & Woollams, 2015; Hsiao et al., 2015). This is not surprising; for content words at least, a word can only appear in more varied contexts if its meaning extends to fit the various contexts (Bolger et al., 2008), potentially making its meaning more nuanced, and hence, more difficult to judge in semantic tasks such as relatedness judgement. This has been the "go-to" explanation to why high-diversity words tend to be disadvantaged in relatedness judgement (e.g., Chapman, 2019; Johns et al., 2016). However, while Simulation 3a does not negate this possibility, it suggests that this disadvantage may also be related to the fact that high-diversity words had fewer opportunities to be integrated into a discourse topic. A high-diversity word is more likely to appear in more topics than a low-diversity word, and this means that when frequency is matched, high-diversity words must be experienced fewer times in each topic than low-diversity words. This leads to the implication that it is more difficult to integrate a high-diversity word into pre-existing long-term memory, and in turn, potentially making their meanings more difficult to judge. In sum, the high-diversity disadvantage in semantic relatedness judgement task (e.g., Experiment 1; Johns et al., 2016) may not only be a consequence of them being more semantically nuanced, but also related to them having fewer opportunities to be integrated into prior knowledge.

Results from Simulation 3b are harder to interpret—the effect of semantic diversity was inconsistent across window sizes. However, this aligns with the null findings from Experiment 2 where semantic diversity did not have a clear effect on relatedness judgement. These computational and behavioral findings, therefore, suggest that anchoring high-diversity novel words can potentially help them integrate into pre-existing knowledge. As a consequence, they might be judged as well as their low-diversity counterparts in semantic relatedness.

The simulations presented in this chapter have several limitations that need to be acknowledged. First, they are clearly some distance from adult word learning, which happens in the context of a large and complex knowledge base. Future work can consider pre-training a neural network on a large corpus before exposing it to the reading materials in Experiments 1 and 2. Second, the current simulations were built on the assumption that target words are better integrated into pre-existing knowledge when the neighborhood relatedness scores are higher. This is a novel approach developed for this thesis. Future work is needed to evaluate the utility of the neighborhood relatedness score in predicting semantic learning and semantic integration.

Chapter 5

Experiment 3

How does semantic diversity influence online processing? An eye-tracking study

1. Introduction

In Experiment 2 (Chapter 3), novel words experienced in the high- (vs. low-) diversity condition were identified faster in speeded recognition and exerted greater top-down influence in pseudoword inferiority. Two key explanations were put forth to explain this processing advantage in word recognition. Reiterating briefly here, the first explanation is based on the Semantic Distinctiveness Model (Jones et al., 2012; 2017). This argues that when people encounter a word, the current context will be compared to the word's contextual history, built from previous encounters; if the current context is distinct from the stored contextual history, the word will be encoded more strongly than if the current context is identical (or highly similar) to previous experience (Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011; Bolger et al., 2008; Johns et al., 2016). As a result of this, words appearing in more diverse contexts might be represented more strongly in the mental lexicon, subsequently supporting efficient word recognition. Relatedly, words that have been encountered in more diverse contexts may become less associated with a specific context, or in other words, they have become more context-independent (Jones et al., 2012). A behavioral marker of this, as others have argued, might be enhanced word identification in isolation (Hsiao et al., 2020) and in new contexts (Pagán et al., 2020). These two accounts, which I refer to as the *encoding account* and the *context-independence account* respectively, may both contribute to the high-diversity advantage seen in word recognition, or it might be that just one of them is sufficient to explain the recognition advantage observed in Experiment 2.[9] However, it is difficult to infer from behavioral measures based on button presses how good these explanations actually are. As the two accounts make different predictions with regard to online processing, one way to evaluate their explanatory power is to monitor eye movement as participants read passages containing words that vary in semantic diversity.

The encoding account predicts that when a word is encountered in an unexpected topic, the word will be encoded more strongly; readers should therefore show longer fixation durations, compared to another word that always appears in one topic (Johns et al.,

[9] Or it might be that both accounts are wrong.

2016; Pagán & Nation, 2019). Support from this account comes from an eye-tracking study by Pagán and Nation (2019). They asked undergraduates to read rare words embedded in four sentences. The sentences either repeated verbatim four times, or varied from presentation to presentation. Reading times were longer for words in the diverse condition than the repeating condition. This suggests that diverse (vs. repeated) contexts could trigger stronger and more effortful encoding. However, this particular study was concerned with contextual repetition (repeated vs. different contexts), instead of semantic diversity (same-themed vs. different-themed); furthermore, Pagán and Nation explicitly told their participants that the reading materials contained unknown words and this might have encouraged more intentional learning strategies. Therefore, it remains to be seen whether contextual variation, as indexed by semantic diversity, influences how strongly a novel word is encoded in real-time during more incidental learning condition.

The context-independence account, on the other hand, predicts that high-diversity words are easier to integrate into a novel context. Therefore, reading times should be shorter for high- (vs. low-) diversity words when they appear in a never-encountered context. Support for this comes from a study with children (Pagán et al., 2020), in which existing words with high semantic diversity values (derived from Hsiao & Nation's (2018) norms) were fixated for less time than frequency-matched low-diversity words when presented in neutral sentences (e.g., *The family looked for the missing candle (low diversity)/camera (high diversity) before their friends arrived*). As noted in Chapter 2, however, it is not entirely clear what corpus-derived semantic diversity captures and whether it is confounded with other lexical variables.

Clearly, there is a need to use a word learning paradigm to investigate the online reading profile of words that vary in contextual history. This will establish initial data to inform both the encoding and the context-independence accounts. While the experiment presented in this chapter cannot decisively pull these two accounts apart, it can shed light on the extent to which they are needed to explain the high-diversity recognition advantage seen in Experiment 2 and correspondingly, in previous experiments that use corpus-based semantic diversity values for existing words (e.g., Hoffman et al., 2013; Hsiao & Nation, 2018; Hsiao et al., 2020).

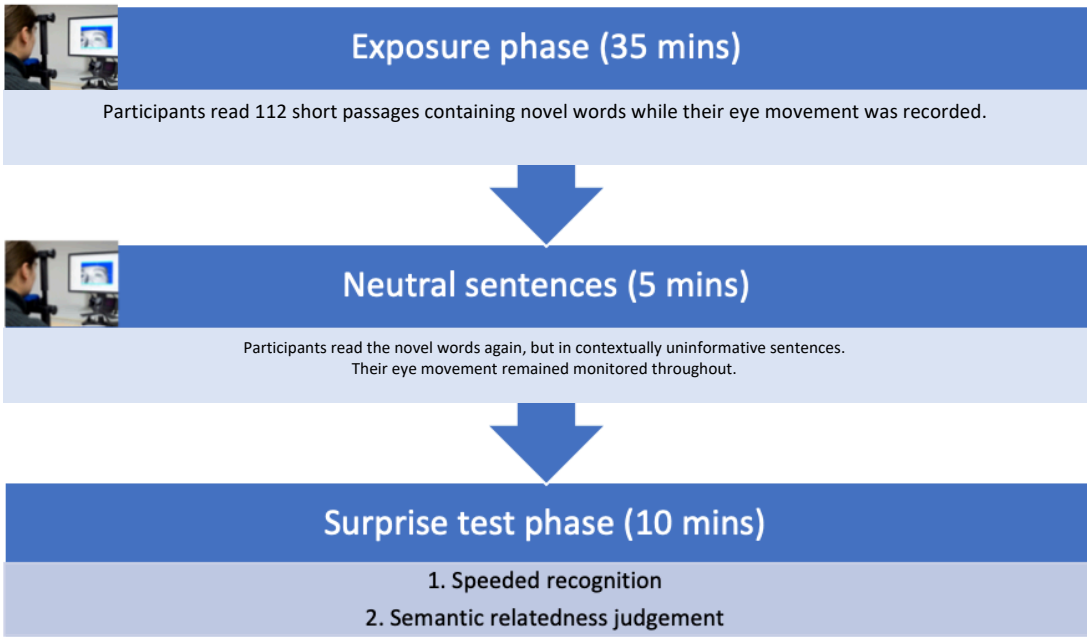
2. Experiment 3

2.1. Overview

Experiment 3 adopted a similar design to Experiment 2, but eye movements were monitored as participants read novel words in naturalistic passages. The experiment comprised three phases (visualized in Figure 5.1): (i) an exposure phase, where variation in semantic diversity for novel words was induced; (ii) a neutral sentence phase, where the participants read the novel words embedded in contextually uninformative sentences; (iii) a surprise test phase, where participants completed two offline tasks measuring their knowledge of the pseudowords: speeded recognition and semantic relatedness judgement.[10] This study was pre-registered ahead of data collection (https://aspredicted.org/blind.php?x=u2p25t)[11].

Figure 5.1

Procedure of Experiment 3.



[10] The pseudoword inferiority task was not administered in Experiment 3 because pilot participants (*N* = 5) commented that the overall experiment was too long and demanding.

[11] In the pre-registration document, there is a typo in question 2: "lower context-independence" should be "*higher* context-independence". In addition, as Experiment 3 was my first ever eye-tracking experiment, I pre-registered an analysis that was not entirely appropriate: An analysis on the spillover regions was pre-registered; however, this was not possible as most spillover regions were not strategically positioned on the computer screen to allow for accurate recording. As a result, eye movement data for the spillover regions could not be meaningfully analyzed, and hence, they are not reported here.

2.2. Method

2.2.1. Participants

Thirty-seven participants (11 males; M_{age} = 18.4; SD_{age} = 0.5) from the University of Oxford took part in the study for partial course credits.[12] All reported to be native speakers of English, with normal or correct-to-normal vision and no history of any language or attentional disorders. Four of the participants were excluded due to tracker loss or equipment failure. All subsequent analyses were therefore based on the remaining 33 participants.

2.2.2. Materials

A total of 16 relatively infrequent English words were chosen to serve as the base words: 12 of them were taken from Experiment 2 (e.g., *mendacious*) and four new ones were added to increase statistical power (i.e., *abstruse*, *comity*, *extirpate*, *fastidious*; see Appendix 1A for full list). Ten short passages ($M_{words/passage}$ = 35.8, $SD_{words/passage}$ = 8.7) were constructed for each base words: Seven sampled one discourse topic (e.g., Donald Trump), while the other three sampled another topic (e.g., a documentary about Michael Jackson). In addition, two neutral sentences were created for each base word ($M_{words/sentence}$ = 9.2, $SD_{words/sentence}$ = 2.1). Care was taken to make the sentence contexts as uninformative as possible so as not to provide any clues to the core meaning of the base words (e.g., *I do not like mendacious people*; see Appendix 5A for full list of neutral sentences).

The 16 base words were then replaced by pronounceable pseudowords (e.g., OSSANIC, LONBUST, APPURATE), taken from Experiment 2. Two sets of 16 pseudowords were chosen; they were matched on a range of lexical properties, including word length, number of orthographic neighbors, and bigram frequency (see Appendix 1B for full set). Each participant encountered pseudowords from one of the two sets, and assignment to pseudoword set was randomized. The unstudied word set served as the foils in the speeded recognition task.

2.2.3. Apparatus

Eye movements were recorded using an SR Research Eyelink 1000 tracker (SR Research Ltd., Ontario, Canada; spatial resolution 0.05°, sampling rate 1,000 Hz). The reading materials were presented on a 15” Dell monitor set at a refresh rate of 60 Hz with a 1,024 × 768

[12] I pre-registered to recruit 25 participants. However, the experiment attracted more interests from undergraduates than I anticipated. I did not analyze the data when 25 participants had completed the study. Instead, I analyzed the data only when all 37 sign-ups had taken part.

resolution and viewing distance of 60 cm. Lowercase black font (Courier New, size 15) on a gray background was used; three characters subtended 1° of visual angle. Reading was binocular but only the right eye was recorded. For unknown reasons, one participant's right eye could not be tracked properly; their left eye was recorded as a result. Throughout the exposure and neutral sentence phases, participants were asked to refrain from blinking as much as possible during reading and to put their chins on a chin rest to minimize head movements.

2.2.4. Procedure

The implementation of semantic diversity largely follows that of Experiment 2 and is summarized in Table 5.1.

Table 5.1

How semantic diversity was implemented in the exposure phase of Experiment 3.

	Exposure phase						
	Anchoring passages				Post-anchoring passages		
	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Passage 6	Passage 7
High-diversity word	Topic A	Topic A	Topic A	Topic A	Topic C	Topic C	Topic C
Low-diversity word	Topic B	Topic B	Topic B	Topic B	Topic B	Topic B	Topic B

A pseudoword in the high-diversity condition was initially experienced in one topic in the first four passages (anchoring passages) and then in another topic in the fifth to seventh passages (post-anchoring passages). On the other hand, a pseudoword in the low-diversity condition was experienced in seven passages all centered on one discourse topic throughout the exposure phase. Each participant read a total of 16 pseudowords, half in the high-diversity condition, the other half in the low-diversity condition, meaning that semantic diversity was manipulated within-participant. Regardless of the assigned condition, each pseudoword was seen a total of seven times in the exposure phase, maintaining frequency across diversity. Assignment to conditions was counterbalanced on the item level.

The experiment started with a calibration procedure, during which participants looked at each of nine fixation points extending horizontally across the top, center, and bottom of the computer monitor. Calibration accuracy was always below 0.33°; otherwise, calibration and validation were repeated. Participants then entered the exposure phase, where they were asked to read short passages silently and for comprehension. As per the previous experiments, participants were naïve to the purposes of the experiment. At the beginning of each trial, drift measurements were taken and calibration repeated when necessary, and participants were asked to blink once in order to minimize blinking during the

trial. Then, a small fixation cross appeared on the top left of the screen. Once fixated, this triggered the appearance of a passage. Participants had unlimited time to read the passage, after which the participants pressed a button on a Microsoft games controller to proceed to the next trial. For half of the trials, the passage was followed by a true-false comprehension question. This was not connected to the novel word but was included to encourage reading for meaning.

All participants first read two practice passages, followed by a total of 112 experimental passages plus 14 filler passages. These 126 passages were equally divided into seven blocks. Each experimental passage contained one target pseudoword, and each pseudoword was read once in each block. The filler passages did not contain any of the pseudowords but had a relatively infrequent word (e.g., *ameliorate*), designed to reduce suspicion that the targets were pseudowords. Block order within the anchoring and post-anchoring phases was randomized, so was the trial order within each block. The exposure phase required approximately 35 minutes.

Immediately after the exposure phase, participants moved into the neutral sentence phase. Here, they read the newly learned pseudowords in neutral sentences while their eye movements continued to be monitored. No additional instructions were given, so from a participant's perspective, the neutral sentence phase was not clearly distinguishable from the exposure phase. Each pseudoword was seen twice, once in two different neutral sentence frames (see Appendix 5A for full list). All participants read the same 32 neutral sentences and order was randomized.

Finally, participants completed two surprise offline tests: speeded recognition and semantic relatedness judgement. The procedure for each was identical to those used in Experiments 1 and 2, with one modification: In speeded recognition, the stimuli were presented for 800 ms, instead of 1800 ms. This reduction was motivated by the fact that each pseudoword was seen more often in the current experiment (7 in exposure phase plus 2 more in neutral sentences) prior to the surprise tests (cf. 6 exposure in Experiment 1 and 8 exposure in Experiment 2).

2.3. Results

Accuracy on the comprehension questions was high, with all participants scoring at least 88% (M = 95.2%, SD = 4.8%). There was no difference in comprehension for high and low diversity passages [M_{High} = 94.9% vs. M_{Low} = 95.3%; $t(32) = 0.87$, $p = .32$].

Based on the pre-registered analysis plan and informed by previous studies (e.g., Chaffin et al., 2001), six dependent measures were extracted from the eye movement record and used to determine if the diversity manipulation influenced the real-time processing of the target pseudowords. These measures are summarized in Table 5.2.

Table 5.2

Summary of the six eye movement measures (alongside their descriptions).

Eye movement measures	Descriptions
(I) First fixation duration	Initial first-pass fixation on the target.
(II) Gaze duration	The sum of all first pass fixations made on the target.
(III) Go past times	The sum of all temporally contiguous fixations until the point of fixation progresses to the region to the right.
(IV) Total reading times	The sum of all fixations made on the target, including regressions.
(V) Regression in	The probability of making a leftward eye movement into a region having already left that region to the right.
(VI) Regression out	The probability of making a leftward eye movement out of a region before leaving that region to the right.

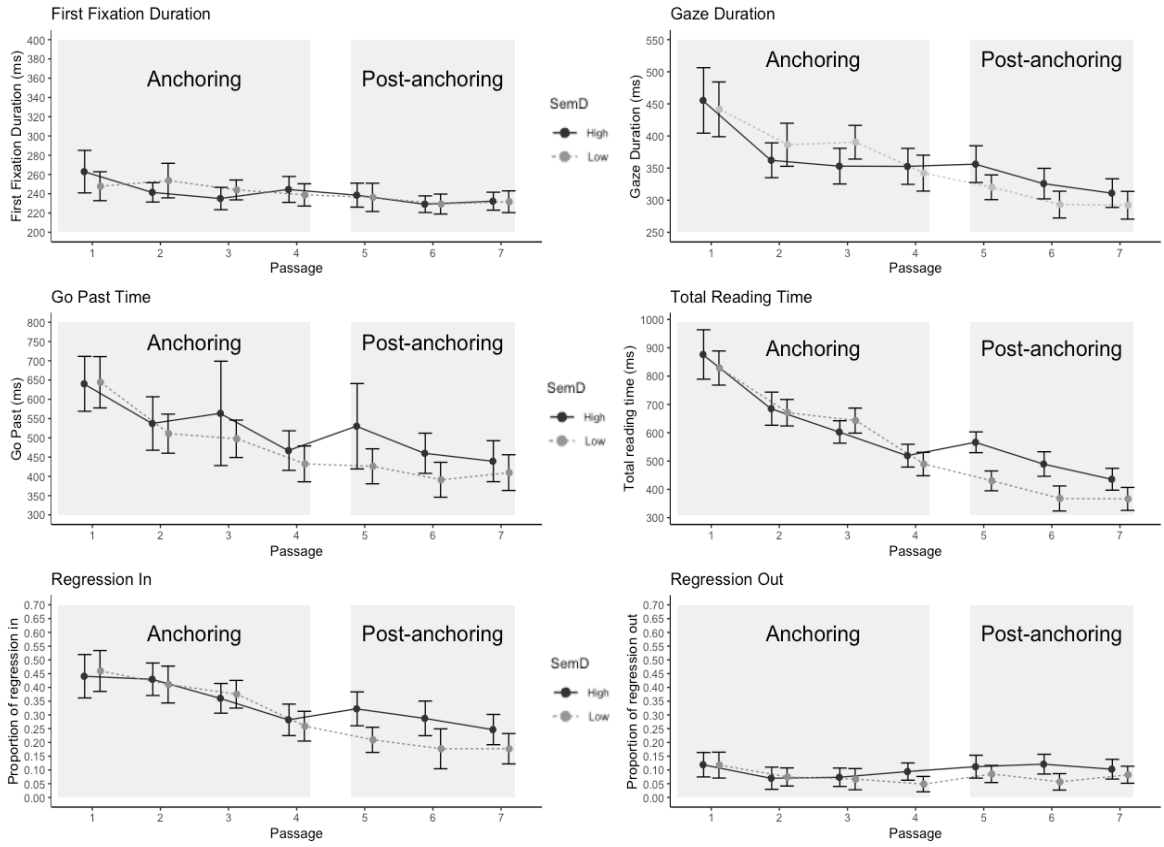
Pre-analysis, the "clean" function in DataViewer (SR Research) was used to trim the data. Fixations shorter than 80 ms and longer than 1200 ms were excluded, and trials that showed blinks or tracker loss on the target word were also discarded (9.3% of all data points).

2.3.1. Exposure phase

Figure 5.2 shows how eye movement changed throughout the exposure phase. Overall, these figures show evidence that the participants learned something about the pseudowords over the course of the exposure phase, with clear decreases in gaze duration, go past time, total reading time and regression in (Blythe, Liversedge, Joseph, White, & Rayner, 2009; Joseph & Nation, 2018; Joseph, Wonnacott, Forbes, & Nation, 2014).

Figure 5.2

Reading times and regression probabilities on target words in the high-diversity and low-diversity conditions across seven exposure in the exposure phase.



Note. Error bars represent 95% within-subject confidence intervals.

All data were analyzed in a mixed-effects environment using the lme4 package (Bates et al., 2015) in R (version 3.6.1; R Core Team, 2019). Generalized linear mixed-effect models were used for the analyses of regressions in and out, while first fixation duration, gaze duration, go-past time, and total reading time were analyzed using linear mixed-effect models. As is typical in eye movement data, the latter four measures were not normally distributed and so were log-transformed. Following the pre-registered analysis plan, two separate sets of analysis were conducted: One examined the effect of semantic diversity within the anchoring phase (passages 1 to 4) while another within the post-anchoring phase (passages 5 to 7).[13] All models included random effects for participants and items and had

[13] The pre-registered analysis plan also included one extra set of analysis (Analysis 2). Results from this set was essentially identical the one reported here, so I chose not to report it here to avoid over-complication.

semantic diversity (high vs. low) as the sole fixed factor, coded using sum contrasts. The computational procedure was the same as the analyses reported in Experiments 1 and 2.

(i) Anchoring phase

Target pseudoword reading times did not differ as a function of semantic diversity for any of the eye movement measures in the anchoring phase (see Tables 5.3 and 5.4 for model summaries). This was as anticipated, given the high- and low- diversity passages were designed to be "the same" in the anchoring passages (i.e., pseudowords in both conditions were experienced in one discourse topic in the anchoring phase).

Table 5.3

Summary of the LME analyses on the target words for first fixation duration, gaze duration, go past time, and total reading time in the anchoring passages.

	First fixation			Gaze duration			Go past time			Total reading time		
	b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
Intercept	5.43	0.03	214.09*	5.77	0.04	128.34*	6.04	0.05	119.55*	6.32	0.06	110.78*
SemD	<0.01	0.01	-0.02	-0.01	0.01	-1.13	0.01	0.01	0.39	<0.01	0.01	0.29

Note. * denotes statistically significant results ($p < .05$).

Table 5.4

Summary of the GLME analyses on the target words for regression in the anchoring passages.

	Regression in			Regression out		
	b	*SE*	*z*	*b*	*SE*	*z*
Intercept	-0.54	0.12	-4.39*	-2.66	0.21	-12.53*
SemD	-0.01	0.05	-0.15	0.11	0.09	1.28

Note. * denotes statistically significant results ($p < .05$).

(ii) Post-anchoring phase

As summarized in Table 5.5, there was a main effect of semantic diversity in gaze duration (*b* = 0.03, *t* = 2.12), go past time (*b* = 0.04, *t* = 2.59), and total reading time (*b* = 0.09, *t* = 6.34), with high-diversity novel words being read for longer. Semantic diversity also showed a significant main effect for regressions in (*b* = 0.30, *t* = 4.48) and out (*b* = 0.23, *t* = 2.49), with readers making more regressions back into and out of the pseudowords in the high-diversity condition (see model summary in Table 5.6).

Table 5.5

Summary of the LME analyses on the target words for first fixation duration, gaze duration, go past time, and total reading time in the post-anchoring passages.

	First fixation			Gaze duration			Go past time			Total reading time		
	b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
Intercept	5.38	0.02	217.10*	5.63	0.04	142.56*	5.86	0.04	130.39*	5.93	0.05	110.91*
SemD	<0.01	<0.01	0.04	0.03	0.01	2.12*	0.04	0.02	2.59*	0.09	0.01	6.34*

Note. * denotes statistically significant results ($p < .05$).

Table 5.6

Summary of the GLME analyses on the target words for regression in the post-anchoring passages.

	Regression in			Regression out		
	b	*SE*	*z*	*b*	*SE*	*z*
Intercept	-1.30	0.16	-8.27*	-2.56	0.22	-11.89*
SemD	0.30	0.06	4.48*	0.23	0.09	2.49*

Note. * denotes statistically significant results ($p < .05$).

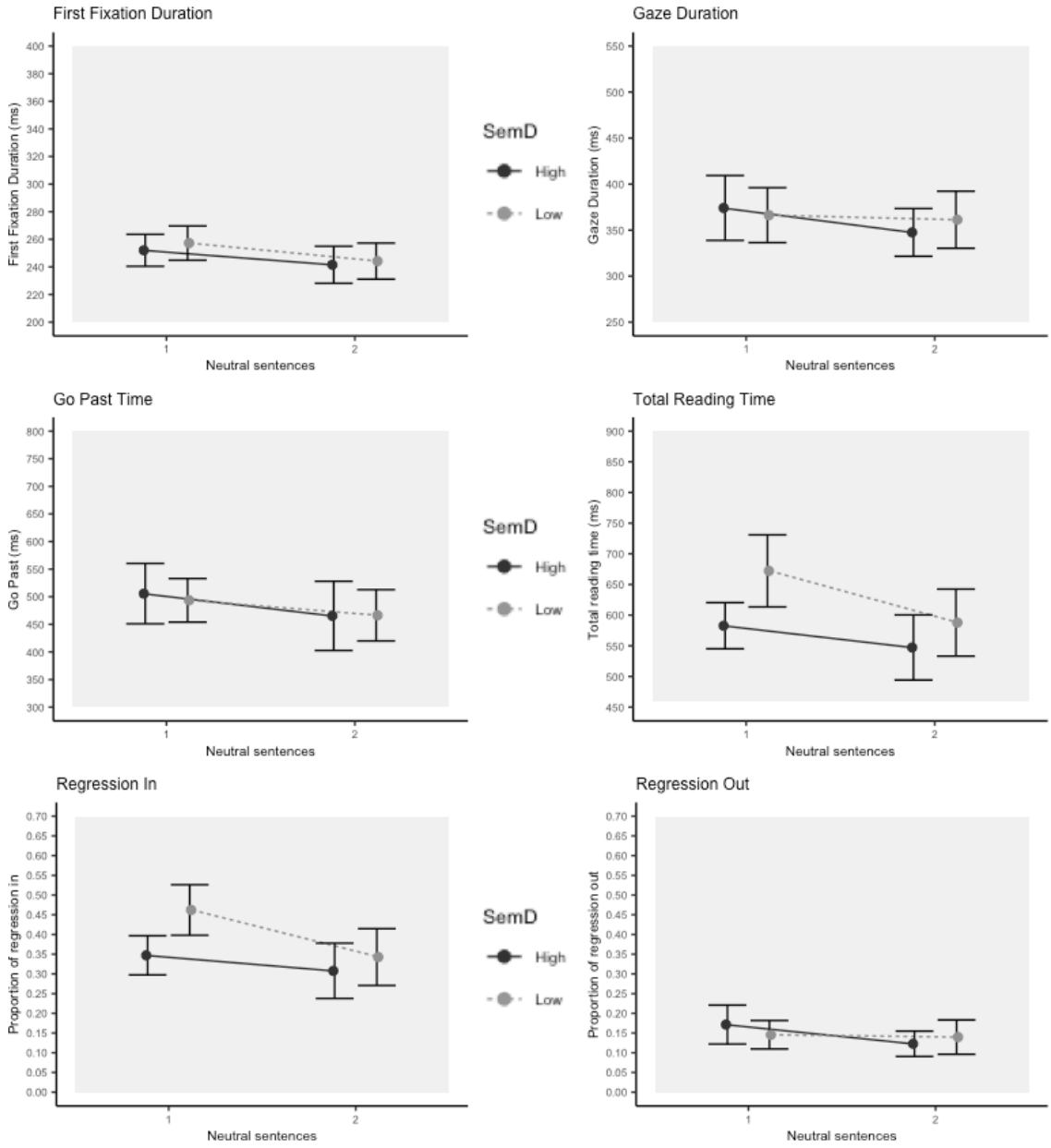
Go past, total reading times, and regressions are typically considered as "late" measures that reflect word-to-text integration rather than word identification (Rayner, 2009). This suggests that in the post-anchoring phase, participants spent more time integrating the high-diversity words into the text following the topic change, relative to the low-diversity novel words that remained in the same topic throughout the exposure phase.

2.3.2. Neutral sentence phase

Reading times for the newly-experienced pseudowords in the neutral sentences are plotted in Figure 5.3.

Figure 5.3

Reading times and regression probabilities on target words in the high-diversity and low-diversity conditions in the neutral sentences.



Note. Error bars represent 95% within-subject confidence intervals.

The mixed-effect models on each of the six measures are summarized in Tables 5.7 and 5.8. These showed a main effect of semantic diversity in total reading time (b = -0.04, t =

-2.17) and regression in (b = -0.16, t = -2.28), with pseudowords in the high-diversity condition receiving *shorter* total reading time and *fewer* regressions. These findings stand in contrast to those from the exposure phase.

Table 5.7

Summary of the LME analyses on the target words for first fixation duration, gaze duration, go past time, and total reading time in the neutral sentences.

	First fixation			Gaze duration			Go past time			Total reading time		
	b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
Intercept	5.44	0.03	198.39*	5.75	0.05	126.27*	5.96	0.06	107.54*	6.21	0.06	100.47*
SemD	-0.01	0.01	-0.61	-0.01	0.02	-0.58	-0.01	0.02	-0.38	-0.04	0.02	-2.17*

Note. * denotes statistically significant results (p < .05).

Table 5.8

Summary of the GLME analyses on the target words for regression in the neutral sentences.

	Regression in			Regression out		
	b	*SE*	*z*	*b*	*SE*	*z*
Intercept	-0.60	0.14	-4.15*	-2.02	0.22	-9.28*
SemD	-0.16	0.07	-2.28*	0.01	0.10	0.11

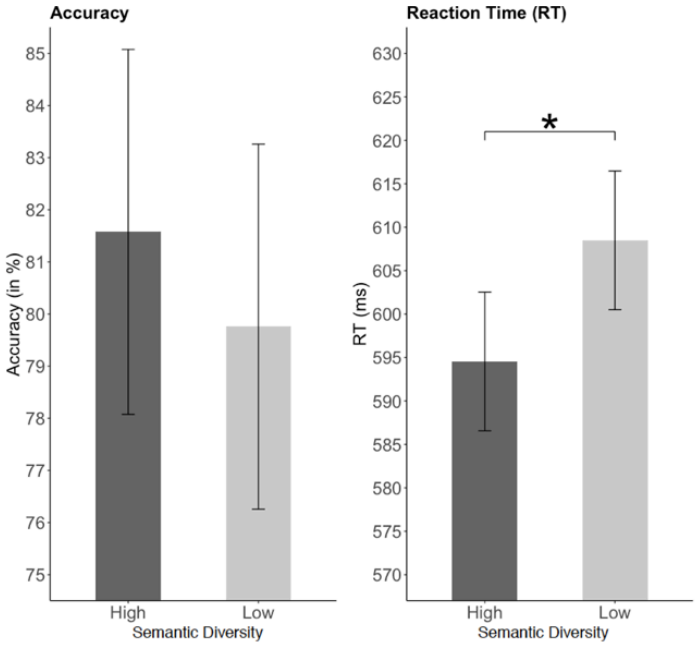
Note. * denotes statistically significant results (p < .05).

2.3.3. Test phase

***Speeded recognition*.** Mean accuracy rates in the two diversity conditions are summarized in the left panel of Figure 5.4. While accuracy is numerically higher in the high-diversity condition (M_{High} = 81.7% vs. M_{Low} = 79.8%), the difference is not statistically significant (z = -0.21, p = .347; see Table 5.9 for model summary). Turning to RT, only correct responses to the targets were analyzed (80.7% of all target trials). Mean RTs in the two diversity conditions are summarized in the right panel of Figure 5.4. Participants were significantly faster at identifying the novel words that were experienced in the high-diversity passages (M = 591ms) than those in the low-diversity passages (M = 607ms, t = -2.83, p = .005), with a mean 16-ms advantage. This replicates the finding from Experiment 2, suggesting that the high-diversity novel words were more accessible than their low-diversity counterparts.

Figure 5.4

Mean accuracy (left) and RTs (right) in the two diversity conditions in speeded recognition, Experiment 3.



Note. Error bars represent 95% within-subject confidence intervals.

Table 5.9

Summary of mixed-effects models examining the effect of semantic diversity on accuracy and RTs in speeded recognition, Experiment 3.

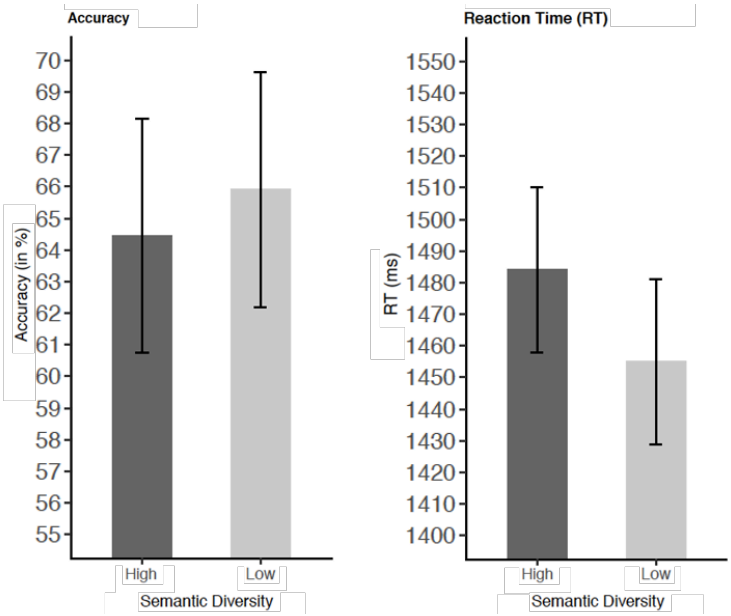
	Accuracy				Reaction Times			
	b	*SE*	*z*	*p*	*b*	*SE*	*t*	*p*
Intercept	1.82	0.23	8.07	<.001*	-169.4	1.96	-86.54	<.001*
SemD	0.11	0.12	0.94	.347	-1.83	0.65	-2.83	.005*

Note. * denotes statistically significant results ($p < .05$). To improve interpretability, estimate and standard error for the RT data were multiplied by 100,000.

Semantic relatedness. Overall accuracy is satisfactory (M = 65.2%, SD = 9.3%) and is significantly above chance [one sample t-test: $t(32) = 9.33$, $p < .001$], indicating that participants had gained some knowledge of the meanings of the newly learned words. Accuracy rates and RTs in the two diversity conditions are plotted in Figure 5.5. Accuracy was similar across the two conditions (M_{High} = 64.6% vs. M_{Low} = 65.9%; $z = 0.82$, $p = .414$, see Table 5.10 for model summary) and RTs to correct trials (65.2% of all trials) were also not significantly different between conditions (M_{High} = 1470 ms vs. M_{Low} = 1449 ms; $t = 1.63$, $p = .102$). These findings mirror those from Experiment 2.

Figure 5.5

Mean accuracy (left) and RTs (right) in the two diversity conditions in semantic relatedness judgement, Experiment 3.



Note. Error bars represent 95% within-subject confidence intervals

Table 5.10

Summary of mixed-effects models examining the effect of semantic diversity on accuracy and RTs in semantic relatedness judgement, Experiment 3.

	Accuracy				Reaction Times			
	b	*SE*	*z*	*p*	*b*	*SE*	*t*	*p*
Intercept	0.65	0.11	5.58	<.001*	726.5	2.11	344.64	<.001*
SemD	-0.04	0.05	-0.82	.414	0.94	0.58	1.63	.102

Note. * denotes statistically significant results ($p < .05$). To improve interpretability, estimate and standard error for the RT data were multiplied by 100.

2.4. Discussion

Using eye-tracking, Experiment 3 investigated how semantic diversity influences the real-time processing of novel words embedded in naturalistic texts. The study began with an exposure phase to induce variation in semantic diversity. Building from Experiment 2, the exposure phase was divided into two blocks: an anchoring block (passages 1 to 4) and a post-anchoring block (passages 5 to 7). A novel word in the low-diversity condition was experienced in one topic throughout while a novel word in the high-diversity condition was experienced first in one topic in the anchoring block and then in another in the post-anchoring block. As to be expected, high- and low- diversity novel words in the anchoring

block produced identical reading behavior: There were no differences in any eye movement measures. However, in the post-anchoring block, novel words in the high-diversity condition were read for longer and received more regressions. This supports the notion that high-diversity contexts could trigger stronger encoding. Following the exposure phase, participants read all of the newly learned words in contextually uninformative sentences. Once again there was a difference in reading behavior as a function of semantic diversity. This time, however, the high-diversity words received shorter reading time and fewer regressions in than low-diversity words. This suggests that the high-diversity words could be integrated into neutral contexts more easily. This provides support for the context-independence account. Finally, the experiment concluded with two offline tasks: speeded recognition and semantic relatedness judgement. Results from these tasks mirror those from Experiment 2: Faster recognition for newly learned words experienced in the high-diversity condition, and no reliable difference between the diversity conditions in relatedness judgement. Below I consider in more detail the two theoretical accounts—the encoding and context-independence accounts.

High-diversity words appeared to be encoded more strongly when the topic changed

When the topic changed in the post-anchoring block, word targets experienced in the high-diversity condition (vs. low-diversity condition) received longer gaze duration, longer go past time, longer total reading time, and more regressions. The latter three measures are often considered as indices of "late" processing (Rayner, 2009), so the current findings suggest that readers spent more time, and perhaps more cognitive effort, integrating the high-diversity words into the context when the discourse topic changed. These findings are similar to Pagán and Nation's (2019) who reported longer viewing times for low-frequency words experienced in four different sentences than those experienced in the same sentence repeated four times. Taken together, these findings lend support to the Semantic Distinctiveness Model (Johns et al., 2014; Jones et al., 2012), which predicts that readers are highly sensitive to a word's contextual environment and that a word will be encoded more strongly when it is encountered in more diverse (vs. repeated or semantically similar) contexts.

Experiment 3 bears some resemblance to Joseph and Nation's (2018) experiment, in which children (aged 10 to 11 years) read unfamiliar low-frequency verbs (e.g., *exacerbate*, *languish*) embedded in 10 different sentence frames that were either high or low in semantic diversity. There were no differences in reading behavior for words experienced in

the two diversity conditions, standing in contrast to the current study. This discrepancy may be due to several factors. First, participants in Joseph and Nation's (2018) experiment were children while those in the current study were young adults. How age, semantic diversity, and prior knowledge influence incidental word learning remains an unexplored question. Second, the high-diversity target words in the current study were given an anchoring opportunity, unlike those in Joseph and Nation's experiment. Third, semantic diversity was operationalized somewhat differently in the two studies, as illustrated in Tables 5.11 and 5.12.

Table 5.11

Reading materials in the low-diversity condition in Joseph and Nation (2018).

Condition: Low Semantic Diversity	**Topic: Animals**	**Target word: *Languish***
1. The animals had languished because they weren't being looked after properly.		
2. The dog languished in her basket for a week before her owner realized she was pregnant.		
3. The farmer claimed that the horse languished after her foal was taken away from her.		
4. When the old lady became unwell, her canary languished as there was nobody to feed		

Table 5.12

Reading materials in the low-diversity condition in Experiment 3.

Condition: Low Semantic Diversity	**Topic: Neil Armstrong**	**Target word: *Diffident***
1. Neil Armstrong, the first man to step foot on the moon, never sought the spotlight. He was regarded as diffident by some, distant by others. What this biography about Armstrong reveals is an astronaut of extreme focus, competence and bravery.		
2. Neil Armstrong's diffident nature made him appear ill at ease at news conferences, and the brevity and occasional evasiveness of his answers led to some unfavorable comments.		
3. Neil Armstrong was regarded as diffident and reserved. He spent two days with his father and never once mentioned that the day after they parted, he was going to be officially named as the first man to set foot on the moon.		
4. During the Cold War, America needed a hero and Armstrong was it. Described as diffident and quiet, he was characteristically self-effacing when interviewed about his role in the Apollo 11 mission.		

In Joseph and Nation (2018), while the sentences in the low-diversity condition were all about animals, each of them was about a different type of animals (e.g., dog, horse, canary). But in Experiment 3, the topics of the low-diversity passages were mostly restricted to one hyponym; that is, if a set of low-diversity passages were about Neil Armstrong, they were only related to Neil Armstrong, not other astronauts like Buzz Aldrin or Helen Sharman. In other words, the degree of semantic overlap within the low-diversity condition in Experiment 3 was considerably higher than that in Joseph and Nation (2018), potentially

explaining why the high- and low-diversity conditions did not differ in any eye movement measures in their study. These contrasting results highlight the need to understand the nature of semantic diversity and agree its definition.

High-diversity words can be integrated into new contexts more easily

Novel words in the high-diversity condition were easier to read when subsequently encountered in neutral sentences, receiving shorter total reading times and fewer regressions in than the low-diversity counterparts. This suggests that the participants found it easier to integrate the newly learned high-diversity words into neutral contexts (Pagán & Nation, 2019; Plummer, Perea, & Rayner, 2014). It might be that the topic switch in the post-anchoring phase helped the representations of the high-diversity words to become less tightly bound with a specific topic and, therefore, more readily available to be integrated into new contexts (Bolger et al., 2008). An alternative but closely related explanation stems from Perfetti's (2007) lexical quality hypothesis, which argues that experiencing a word in diverse contexts may help refine and boost the quality of its lexical-semantic representation, and this should, in turn, make the word easier to process in never-encountered contexts. The data from Experiment 3 do not allow the differentiation of these two possibilities (Pagán & Nation, 2019), but currently, "lexical quality" and "context-independence" are underspecified notions in the extant literature (Joseph & Nation, 2018), so it is unclear whether the difference between them is just a matter of terminology. Future empirical work is needed to determine the extent to which these constructs can be teased apart. To wrap up, the eye-tracking data from Experiment 3 suggest that the recognition advantage associated with high-diversity words might be related to these words being encoded more strongly during reading and them having grown to be more context-independent. In other words, both the encoding and context-independence accounts may be needed to explain the recognition advantage associated with high-diversity words (e.g., Experiment 2; Hoffman et al., 2013; Hsiao & Nation, 2018). However, future work is needed to investigate how these accounts relate to each other.

Results from the offline tasks replicated those from Experiment 2

The two offline tasks, speeded recognition and semantic relatedness judgement, replicated the results from Experiment 2. In speeded recognition, target words experienced in the high-diversity condition were recognized faster than those seen in the low-diversity condition. The recognition advantage, although small at 16 ms, is similar in magnitude to the one

reported in Experiment 2 (21 ms) and Johns et al. (2016; 26 ms). Note that the stimulus presentation time was reduced from 1800 ms to 800 ms here. This highlights the robustness of the high-diversity advantage in visual recognition. As for accuracy, overall performance was significantly poorer in this experiment (*M* = 80.6%, *SD* = 14.7%) than in Experiment 2 (*M* = 92.0%, *SD* = 6.5%). It might be that the reduction in stimulus presentation time increased task difficulty. In neither experiment was there a significant effect of semantic diversity on accuracy. Performance was heading towards ceiling in Experiment 2, but that was not the case here in Experiment 3. However, Johns et al. (2016) did find a significant effect of diversity on recognition accuracy, with high-diversity words being recognized more accurately. One explanation is that in Experiment 3, the novel words were seen nine times prior to speeded recognition, while those in Johns et al.'s experiment were seen five times only. Plausibly, the effect of semantic diversity on recognition accuracy is influenced by the frequency of exposure such that an effect may be more likely to be detected when a word is less frequent. This possibility is line with the finding that in lexical decision, corpus-derived semantic diversity for existing words shows a larger effect in low than high frequency words, for both accuracy and RT (e.g., Cevoli et al., 2020; Hsiao et al., 2020).

Finally, turning to relatedness judgement, Experiment 1 (Chapter 2) and previous work using corpus-derived semantic diversity (e.g., Hoffman & Woollams, 2015; Hsiao et al., 2020) converge in showing that high-diversity words are more difficult to process in tasks that require specific semantic details. In contrast, Experiment 3 found no effect of semantic diversity on the relatedness judgement task. This mirrors the null finding seen in Experiment 2 (Chapter 3) and the inconsistent diversity effect observed in Simulation 3b (Chapter 4). As described in Experiment 2 and Simulation 3, the lack of a diversity effect in relatedness judgement may be attributed to two factors: First, words in the high-diversity condition were experienced in two topics only, as opposed to six topics in Experiment 1; therefore, the meanings of the high-diversity words might not have grown nuanced enough to interfere with semantic judgement. Second, the high-diversity words in Experiment 3 (vs. those in Experiment 1) were experienced in each discourse topic more frequently, potentially making them better integrated into pre-existing knowledge.

Chapter 6

0. Foreword

Experiment 3 was completed in January 2020. Soon after, COVID-19 came in like a wrecking ball, messing everyone's research plans. The national lockdown from March to June 2020 inevitably prevented me from running another lab-based/eye-tracking study. Experiment 4, therefore, recruited participants online via Prolific. One major concern at the time was that participants might use a dictionary to look up the novel words. Rather than trying to capture incidental learning, therefore, Experiment 4 shifted the focus to more intentional learning, where participants were explicitly told to learn the novel words in the reading materials.

Experiment 4

Can explicit definitions serve to anchor?

1. Introduction

Experiment 1 and Simulation 1 demonstrated that experiencing novel words in a repeated and familiar topic can facilitate initial learning more than experiencing them in a diverse set of familiar topics. This finding was attributed (at least partially) to anchoring, the process by which novel words are secured onto a pre-existing concept in long-term memory. As demonstrated in Experiments 2 and 3, novel words that have been anchored are better placed to take advantage of contextual variation. Results from these behavioral experiments converged with Simulations 1 and 2 to show the utility of anchoring in the initial stages of word learning in adults.

In Experiments 1 to 3, an anchoring opportunity involved presenting novel words in a familiar *topic* repeatedly. In Experiment 4, I asked whether experiencing novel words alongside a familiar *definition* will serve to anchor the novel words. As discussed in Chapter 1 (section 2.2.2), providing age-appropriate definitions as glosses is a common practice in second-language learning textbooks, and there is evidence that this can improve word learning outcome (e.g., Paribakht & Wesche, 1996; Watanabe, 1997). Here, I hypothesize that exposing adult readers to novel words repeatedly with a definition composed of a familiar concept would allow the novel words to be anchored onto the concept's representation in long-term memory, achieving a comparable effect as experiencing novel words in a familiar topic.

2. Overview

Following Experiments 1 to 3, young adults experienced novel words in single-themed passages (low-diversity) or multi-themed passages (high-diversity). Unlike the previous experiments, participants in the current study were informed at the start of the experiment that they would read passages containing novel words and that their knowledge of those words would be subsequently tested. Another major difference in Experiment 4 and is that it employed a 2 x 2 mixed design. The within-subject variable, as per the previous experiments, was semantic diversity (high vs. low). A new addition is a between-subject variable referred to as *group allocation*: Half of the participants were assigned to the *with-definition* group where the word targets were presented alongside a definition composed of a familiar concept before the words were read in a passage. The other half of the participants were assigned to the *without-definition* group where the word targets were presented in isolation before they were experienced in a passage (i.e., the meanings had to be inferred from the contexts). I hypothesized that in the with-definition group, a high-diversity advantage would emerge immediately after exposure in visual recognition and in tasks that tap the degree of context-independence. This is because the provision of a familiar definition might serve to anchor the target words, enabling them to be better placed to take advantage of contextual variation. I also hypothesized that given the low-diversity advantage reported in Experiment 1, the without-definition group would show a low-diversity advantage in both visual recognition and relatedness judgement. In other words, the effect of semantic diversity on visual recognition was predicted to be opposite across the two group allocations: a high-diversity advantage in the with-definition group, but a low-diversity advantage in the without-definition group.

A final addition to Experiment 4 is a follow-up session that was administered approximately 24 hours after the exposure phase. This was added to investigate whether the effect of semantic diversity on lexical processing changes over a period of offline consolidation (e.g., James, et al., 2019). This study was pre-registered ahead of data collection (https://aspredicted.org/blind.php?x=bh6cq9).

2.1. Method

2.1.1. Participants

One hundred and five adults recruited from Prolific (www.prolific.ac) (24 males; M_{age} = 27.8; SD_{age} = 4.18) participated in this experiment for payment (£7/hour). All were native English speakers with normal or corrected-to-normal eyesight, and none reported history of any

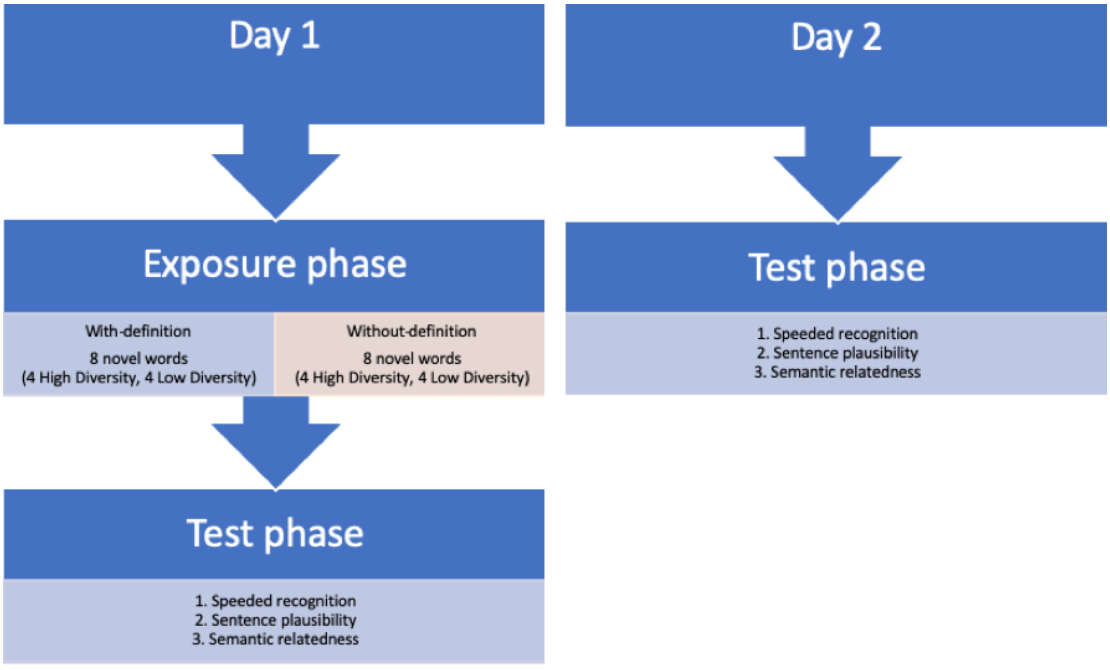
developmental disorders. Five participants were excluded because they failed >20% of the attention checks. All analyses on the Day 1 post-tests, unless otherwise specified, are based on the remaining 100 participants. As for the delayed session on Day 2, 80 of the 100 participants took part in it (Note: the five participants who failed the attention checks on Day 1 were not invited back).

2.1.2. Materials and procedure

This behavioral study comprised two phases (visualized in Figure 6.1): (i) an exposure phase where variation in a target word's semantic diversity was induced, and (ii) a test phase that measured the knowledge and processing of the newly learned words. Participants completed the test phase twice, once immediately after the exposure phase (Day 1) and another 24 hours later (Day 2). All the tasks were programmed and run using Gorilla (www.gorilla.sc). The sessions on Day 1 and 2 required approximately 40 minutes and 15 minutes, respectively.

Figure 6.1

Procedure of Experiment 4.



(i) Exposure phase

The same 16 base words (e.g., *mendacious*, *imbroglio*) from Experiment 3 were used here. Eight passages were constructed for each base word: Four sampled the same discourse topic and formed the low semantic diversity condition; another four spanned four distinct topics and these formed the high semantic diversity condition. The passages were modified from those developed for Experiment 3 and were built on topics deemed familiar to an average British citizen (e.g., Donald Trump, Kate Middleton). The base words were then replaced by the sets of pronounceable pseudowords used in Experiment 3 (e.g., OSSANIC, LONBUST; see Appendix 1B).

Each participant encountered a total of eight pseudowords, four in the high-diversity condition and four in the low-diversity condition. In other words, semantic diversity was manipulated within-subject, as per the previous experiments. Assignment to the semantic diversity condition was fully counterbalanced on the item level.

Half of the participants were randomly allocated to the with-definition group, where they were presented with each target word alongside a definition composed of a familiar concept (e.g., *ossanic* = not telling the truth; see Appendix 6A for the full list of definitions) before reading a passage containing that word.[14] For the other half of the participants, they were assigned to the without-definition group, where a target word was presented alone before it was read in a passage.

A reading trial started with a 400-ms fixation cross at the center of the screen, followed by the presentation of a target word, with or without a definition (depending on which group a participant was allocated to). This stayed on the screen for four seconds and was replaced by a passage containing that target word. Participants were asked to rate how comprehensible the passage was on a continuous scale of 1 (No understanding) to 7 (Perfect understanding). They had unlimited time to read and rate each passage. After the rating was submitted, half of the passages were followed by a simple true-false question, which served to encourage reading for meaning. All the questions were not related to the target words, and pilot testing confirmed that question difficulty was equivalent across diversity conditions.

[14] In language textbooks, it is common for the definitions of difficult words to appear at the text margin. However, if the definitions were shown at the text margin here, participants could ignore the definition or some definitions would receive more attention than the others. To tackle these potential pitfalls, I decided to show the definitions of all the target words for a fixed amount of time (4 seconds) before a passage containing that word was displayed. This gives some assurance that each definition received a similar amount of attention.

Each participant read a total of 32 (8 words x 4 passages) unique passages, divided equally into four blocks. Block order was randomized, as was the trial order within each block. In summary, each participant experienced four pseudowords in the low-diversity condition and four in the high-diversity condition. Regardless of the assigned condition, each pseudoword was seen four times with a definition (in the with-definition group) or in isolation (in the without-definition group), plus four more times in a passage.

(ii) Test phase (Day 1)

Immediately following the exposure phase, participants completed three post-tests in a fixed order, namely speeded recognition, sentence plausibility judgement, and semantic relatedness judgement.

Speeded recognition. As in the previous experiments, participants judged, as quickly and as accurately as possible, whether a stimulus was one of the eight words they learned in the exposure phase. Almost all parameters were identical to the previous experiments. There were however two modifications. First, given participants were explicitly told to learn the word targets, the stimulus presentation time was reduced from 800ms (Experiment 3) to 300ms. This helped increase task demands. Second, the number of blocks was increased from three to four, meaning that the participants completed a total of 64 trials (8 newly learned words x 4 + 8 foils x 4). The increase in block number was due to a concern over reduced statistical power (i.e., in the current experiment, the number of items per condition was four, compared to eight in Experiment 3).

Plausibility judgement. Participants were presented with the neutral sentences developed for Experiment 3 (see Appendix 5A) and had to indicate, as quickly and as accurately as possible, whether or not a sentence made sense. Each sentence contained one of the eight newly learned pseudowords. Four sentences were created for each pseudoword: two where the pseudoword fitted the sentence frame [Yes trial; e.g., *You need to be flematic* (= fastidious) *if you wish to succeed*.], another two where the sentences made no sense [No trial; e.g., *The problems have remained flematic* (= fastidious) *for several years*.].

Each trial started with a fixation cross at the center of the screen for 500 ms, followed by the presentation of a sentence. Participants had five seconds to make a judgement. Participants pressed the “Z” key for “Yes, this sentence makes sense” and the

"M" key for "No, this sentence doesn't make sense". The next trial started automatically after five seconds or once a response was recorded. No feedback was given. Each participant completed a total of 32 trials (8 words x 4 sentences), divided into two blocks. Block order as well as trial order were randomized. The sentence frames were the same for all the participants, but the assignment of pseudowords to sentence frames depended on which set of pseudowords a participant previously learned in the exposure phase. This judgement task started with two practice trials (e.g., *She suffered from an abduction about spider*).

Semantic relatedness. As per Experiments 1 to 3, participants judged, as quickly and as accurately as possible, whether a newly learned word (e.g., *ossanic*) was semantically related to a familiar word (e.g., *dishonest*). All parameters were identical to those in the previous experiments.

(iii) Delayed test phase (Day 2)

Approximately 24 hours following the exposure phase, the participants received an invitation from Prolific to complete a follow-up session. They were given a two-hour time window to start the session. Eighty of the 100 participants took part in this delayed test phase, which was identical to the one administered on Day 1.

2.2. Results

(i) Exposure phase

Performance on the comprehension questions was good (M = 90.4%, SD = 8.06%), indicating that the passages were read for meaning. Comprehension was equivalent across both groups of participants [$t(98.68) = -0.52$, $p = .605$] and across the two semantic diversity conditions [$t(99) = 0.25$, $p = .80$].

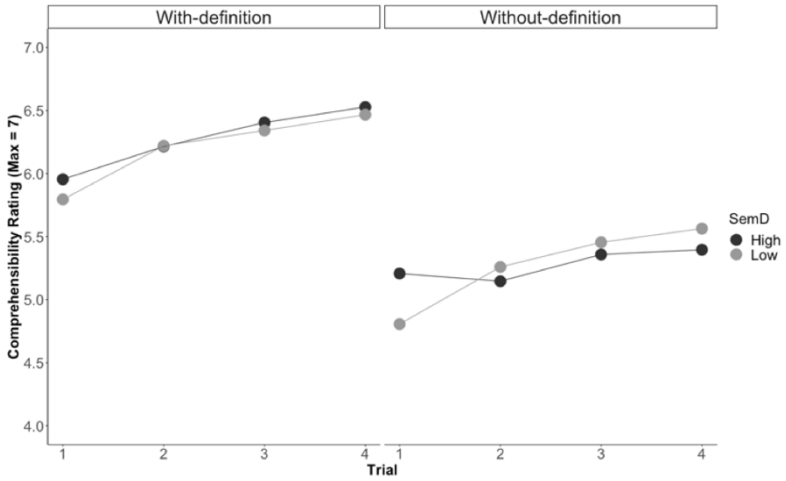
On average, participants spent 17.1 sec (SD = 10.2 sec) to read one passage. Reading times tended to be faster in the with-definition group (M = 16.1 sec, SD = 5.1 sec) than in the without-definition group (M = 17.9 sec, SD = 5.8 sec), although this difference was not statistically significant [$t(100.52) = -1.68$, $p = .095$]. As for the reading times across the diversity conditions, participants spent significantly more time on the high-diversity passages (M = 17.4 sec, SD = 5.6 sec) than on the low-diversity passages (M = 16.7 sec, SD = 5.7 sec), $t(99) = 2.48$, $p = .015$, mirroring the pattern seen in Experiment 1.

Turning to the comprehensibility rating (summarized by group allocations and diversity conditions in Figure 6.2), participants in the with-definition group rated the

passages as more subjectively comprehensible [M_{With} = 6.24 vs. $M_{Without}$ = 5.27; t(3165.1) = 20.47, p < .001], even though the passages were identical for both groups. The effect of semantic diversity on comprehensibility rating was non-significant overall [M_{High} = 5.76 vs. M_{Low} = 5.72; t(1639)= 1.08, p = .279] as well as within each group allocation (ts < 1.6, ps > .11).

Figure 6.2

Comprehensibility ratings for high and low diversity passages across trials and group allocations in Experiment 4



(ii) Test phase

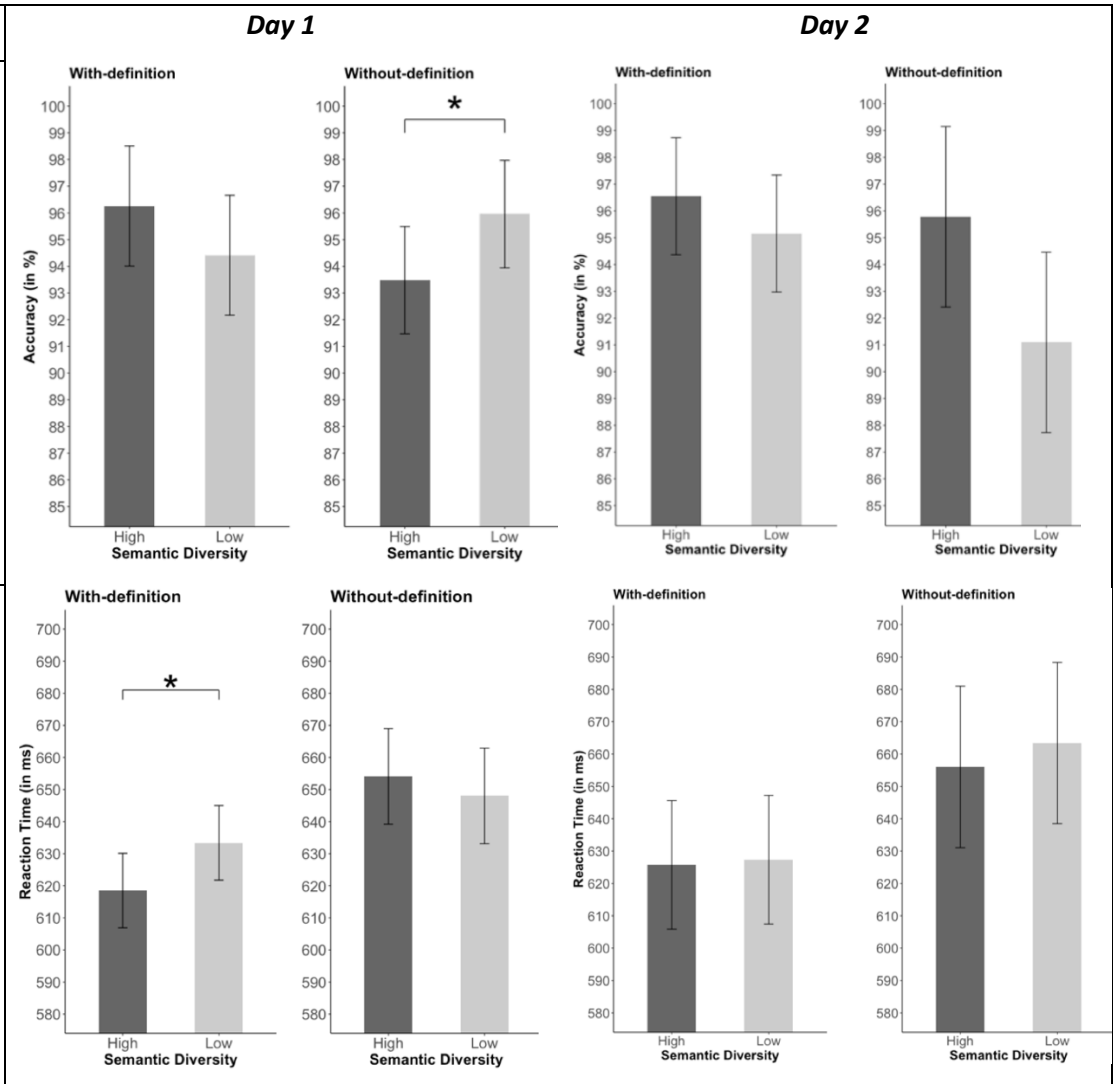
Analysis approach. As per the previous experiments, all data were analyzed in a mixed-effects environment using the lme4 package (version 1.1.17; Bates et al., 2015) in R (R Core Team, 2019). Following the pre-registered analysis plan, all mixed-effects models had two fixed effects, namely semantic diversity (high vs. low) and group allocation (with-definition vs. without-definition), as well as participants and items as random effects. Both of the fixed effects were effect coded. The computational procedures followed those in the previous analyses. Note that two separate sets of analyses were conducted here: one for Day 1, another for Day 2. Day could have been treated as a fixed effect within a single analysis; however, (i) no prior prediction was made with regard to the effect of day (i.e., not pre-registered) and (ii) the drop-out rate was high at 20%. Adding day as a fixed effect will reduce statistical power as participants who did not take part in the session on Day 2 will be discarded from the analysis.

2.2.1. Day 1

Speeded recognition. Data from two participants failed to record properly, meaning these analyses are based on 98 participants. Following the pre-registered analysis plan, trials whose RTs were 2.5 SDs below/above the individual's mean RT were discarded (4.9% of all trials). All analyses were based on the "yes" trials. For RTs, only correct responses were analyzed (94.9% of all "yes" trials) and RTs were subjected to an inverse transformation before analysis. Mean accuracy and mean RTs in the two diversity conditions across the two groups are summarized in the left column in Figure 6.3.

Figure 6.3

Mean accuracy (top) and RTs (bottom) in the two diversity conditions by group allocation and day in speeded recognition, Experiment 4



Note. All error bars represent 95% within-subject confidence intervals.

Accuracy Neither the main effect of semantic diversity (z = -0.48, p = .633) nor group allocation (z = -0.85, p = .396) were statistically significant (see Table 6.1 for model summary). Their interaction, however, was significant (z = -2.88, p = .004). This was followed up by a simple effects analysis that compared the effects of semantic diversity within each group. In the with-definition group, mean accuracy rate was numerically higher in the high-diversity condition (M_{High} = 96.3% vs M_{Low} = 94.4%), but this was not statistically significant (z = 0.20, p = .842). In contrast, in the without-definition group, accuracy was significantly higher in the low-diversity condition (M = 95.9%) than in the high-diversity condition (M = 93.5%), z = 1.97, p = .0495.

Reaction Time There were no effects of semantic diversity (t = -0.90 , p = .373; see Table 6.2) or group allocation (t = 0.90 , p = .372), but there was a significant interaction (t = 2.25, p =.027). This was unpacked by a simple effects analysis, where the high and low diversity conditions were compared within each group allocation. In the with-definition group, pseudowords experienced in the high-diversity condition (M = 634 ms) were responded to faster than those in the low-diversity condition (M = 618 ms), and this was statistically significant (t = -2.24 , p = .030). In contrast, there was no significant difference between the two diversity conditions in the without-definition group (M_{High} = 654 ms vs. M_{Low} = 647 ms; t = 0.90, p = .375).

Table 6.1

*Summary of the GLME models examining the effects of semantic diversity and group allocation on **accuracy** in each of the post-tests, **Day 1** of Experiment 4.*

	Estimate	Standard Error	z value	p value
(d) Speeded recognition				
Intercept	3.29	0.17	19.2	<.001*
SemD	-0.04	0.09	-0.48	.633
Group Allocation	-0.10	0.12	-0.85	.396
SemD x Group	-0.25	0.09	-2.88	.004*
(e) Sentence plausibility				
Intercept	0.80	0.13	6.13	<.001*
SemD	0.10	0.04	2.47	.013*
Group Allocation	-0.15	0.05	-3.01	.003*
SemD x Group	-0.11	0.04	-2.61	.009*
(f) Semantic relatedness				
Intercept	1.05	0.12	9.05	<.001*
SemD	-0.07	0.04	-1.74	.082
Group Allocation	-0.15	0.06	-2.35	.019*
SemD x Group	-0.06	0.04	-1.36	.174

Table 6.2

*Summary of the LME models examining the effects of semantic diversity and group allocation on **RTs** in each of the post-tests, **Day 1** of Experiment 4.*

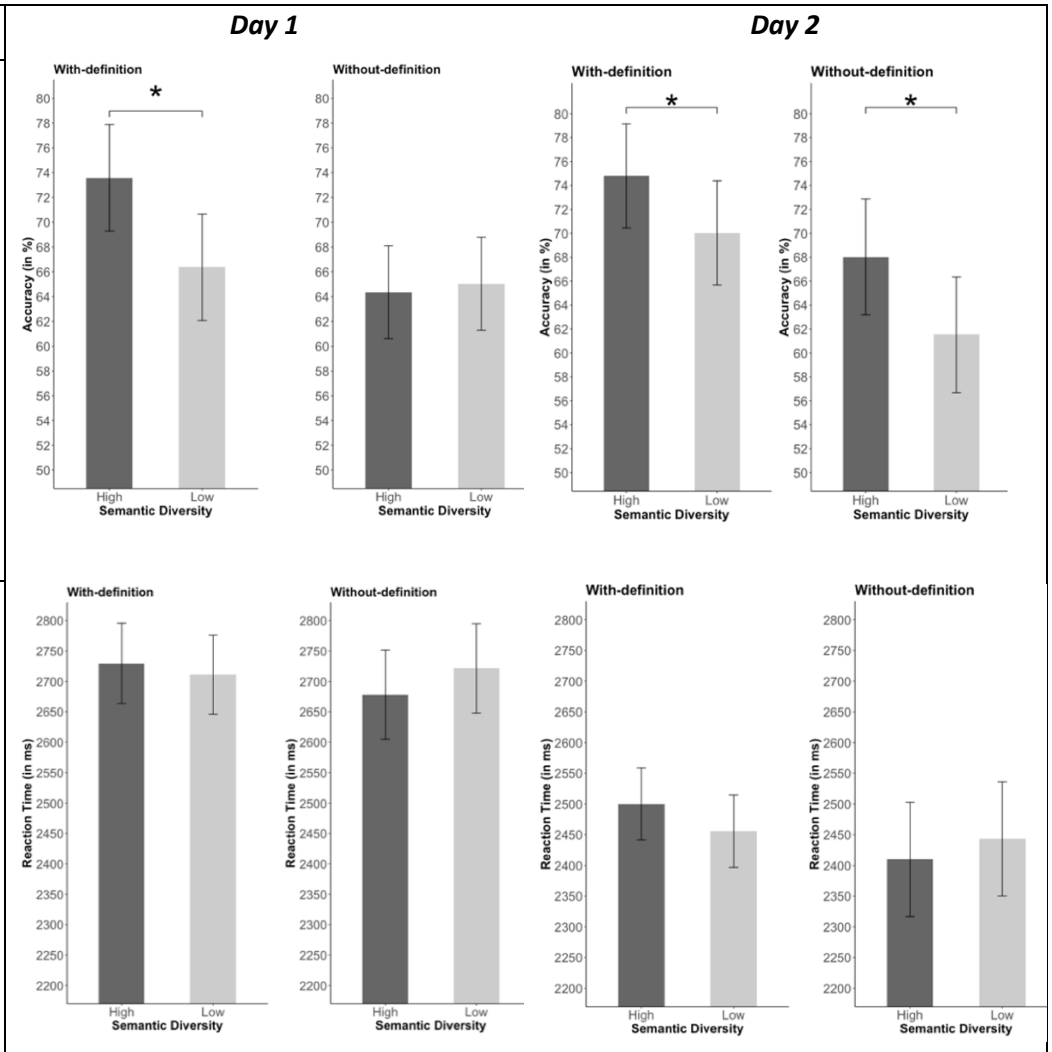
	Estimate	Standard Error	t value	p value
(a) Speeded recognition				
Intercept	-164.4	2.50	-65.67	<.001*
SemD	-0.48	0.54	-0.90	.373
Group Allocation	1.90	2.12	0.90	.372
SemD x Group	1.21	0.54	2.25	.027*
(b) Sentence plausibility				
Intercept	340.9	1.08	314.76	<.001
SemD	-0.15	0.28	-0.55	.584
Group Allocation	-0.24	0.95	-0.25	.800
SemD x Group	-0.27	0.28	-0.94	.345
(c) Semantic relatedness				
Intercept	318.10	1.12	283.27	<.001*
SemD	0.44	0.24	1.85	.064
Group Allocation	0.31	0.76	0.40	.688
SemD x Group	-6.16	2.38	-2.60	.010*

Note. To improve interpretability, estimate and standard error were multiplied by 100,000 in speeded recognition and by 100 in both sentence plausibility and semantic relatedness.

Sentence plausibility. Trials whose RTs were 2.5 SDs below/above the individual's mean RT were discarded (7.3% of all trials). Overall accuracy was satisfactory (*M* = 67.7%, *SD* = 11.6%) and was significantly above chance [one-sample t-test: $t(99) = 15.0$, $p < .001$], showing that participants had acquired some semantic information about the target words. For the analysis on RTs, only correct responses were analyzed (67.7% of all trials), and RTs were subjected to a log transformation before analysis. Mean accuracy rates and mean RTs in the two diversity conditions across the two groups are summarized in the left column of Figure 6.4.

Figure 6.4

Mean accuracy (top) and RTs (bottom) in the two diversity conditions by group allocation and day in sentence plausibility, Experiment 4.



Note. All error bars represent 95% within-subject confidence intervals.

Accuracy The main effects for both semantic diversity (z = 2.47, p = .013) and group allocation (z = -3.01, p = .003) were significant, but they were qualified by a significant interaction (z = -2.61, p = .009). Therefore, a simple effects analysis was conducted. In the with-definition group, pseudowords experienced in the high-diversity condition were more likely to be correctly judged on whether they fitted a neutral sentence frame (M_{High} = 74.7% vs. M_{Low} = 67.0%; t = -3.42, p < .001). In the without-definition group, however, there was no reliable difference between the high and low diversity conditions (M_{High} = 64.5% vs. M_{Low} = 64.9%; t = 0.11, p = .909).

Reaction Time There were no effects of semantic diversity (t = -0.55, p = .584) or group allocation (t = -0.25, p = .80) and no interaction between two (t = -0.94, p = .345).

Semantic relatedness. Responses from one participant failed to record properly, so the analyses here are based on 99 participants. Trials whose RTs were 2.5 SDs below/above the individual's mean RT were discarded (6.3% of all trials). Overall accuracy was good (M = 72.2%, SD = 12.5%) and above chance [one-sample t-test: t(98) = 17.64, p < .001]. For the RT analysis, only correct responses were analyzed (72.2% of all trials), and RTs were subjected to a log transformation before analysis. Mean accuracy and mean RTs in the two diversity conditions across groups are summarized in the left column in Figure 6.5.

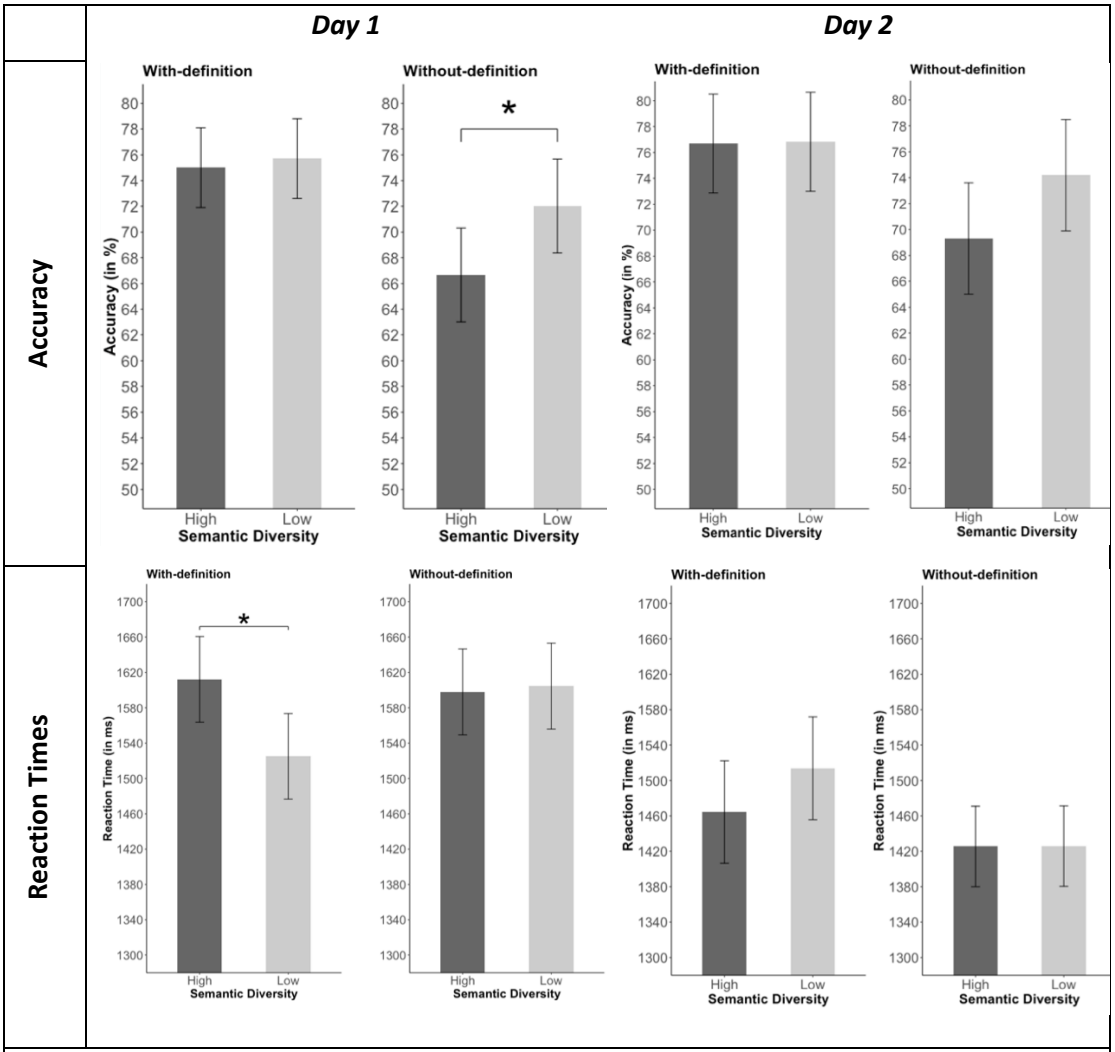
Accuracy There was a significant main effect of group allocation (z = -2.35, p = .019) such that participants who learned the novel words with a definition were more accurate than those who didn't (M_{with} = 75.4% vs. $M_{without}$ = 69.8%). Neither semantic diversity (z = -1.74, p = .082) nor the interaction (z = 1.36, p = .174) were significant. Despite the insignificant interaction, I explored the effect of semantic diversity within each group, motivated by the results from Experiment 1 and Johns et al. (2016). The exploratory analysis revealed that in the without-definition group, judgement was more accurate in the low-diversity condition (M_{Low} = 72.4% vs. M_{High} = 67.1%; z = -2.32, p = .020), in line with the finding from the same task in Experiment 1. In contrast, performance across the two diversity conditions was almost identical in the with-definition group (M_{High} = 75.3% vs. M_{Low} = 75.5%; t = 0.36, p = .719).

Reaction Time There were no effects of semantic diversity (t = 1.85, p = .064) or group allocation (t = 0.40, p = .688), but there was a significant interaction (z = -2.60, p =.010). The

effect of semantic diversity was therefore examined within each group. In the with-definition group, novel words in the low-diversity condition (*M* = 1528 ms) elicited faster judgement than those in the high-diversity condition (*M* = 1614 ms), and this was statistically significant (*t* = 3.18, *p* = .002). In the without-definition group, there was no significant difference between the two diversity conditions (M_{High} = 1589 ms vs. M_{Low} = 1604 ms; *t* = -0.46, *p* = .645).

Figure 6.5

Mean accuracy (top) and RTs (bottom) in the two diversity conditions by group allocation and day in semantic relatedness, Experiment 4.



Note. All error bars represent 95% within-subject confidence intervals.

2.2.2. Day 2

All the analyses in this section followed the same analysis procedure as for Day 1 and are based on data from 80 participants.

Speeded Recognition (Day 2). Mean accuracy and mean RTs in the two diversity conditions across the two groups are summarized in the right column of Figure 6.3.

Accuracy There were no effects of semantic diversity (z = 0.38, p = .698) or group allocation (z = -1.60, p = .109), and their interaction was also not statistically significant (z = 1.34, p = .181) (see Table 6.3 for model summary).

Reaction Time Neither semantic diversity (t = -0.40, p = .693) nor group allocation (t = 1.33, p = .186) had an effect on RTs and their interaction was also not statistically significant (t = 0.12, p = .903) (see Table 6.4 for model summary).

Table 6.3

*Summary of the GLME models examining the effects of semantic diversity and group allocation on **accuracy** in each of the post-tests, **Day 2** of Experiment 4.*

	Estimate	Standard Error	z value	p value
(a) Speeded recognition				
Intercept	3.77	0.23	16.49	<.001*
SemD	0.10	0.25	0.38	.698
Group Allocation	-0.26	0.16	-1.60	.109
Sem D x Group	0.14	0.15	0.97	.334
(b) Sentence plausibility				
Intercept	0.84	0.12	6.75	<.001*
SemD	0.13	0.05	2.79	.005*
Group Allocation	-0.19	0.05	-3.43	<.001*
SemD x Group	0.01	0.05	0.32	.748
(c) Semantic relatedness				
Intercept	1.19	0.14	8.49	<.001*
SemD	-0.05	0.06	-0.96	.337
Group Allocation	-0.15	0.08	-1.88	.060
SemD x Group	-0.06	0.06	-1.05	.292

Table 6.4

*Summary of the LME models examining the effects of semantic diversity and group allocation on **RTs** in each of the post-tests, **Day 2** of Experiment 4.*

	Estimate	Standard Error	z value	p value
(a) Speeded recognition				
Intercept	-168.4	2.70	-60.98	<.001*
SemD	-0.28	0.71	-0.40	.693
Group Allocation	3.17	2.38	1.33	.186
SemD x Group	0.09	0.71	0.12	.903
(b) Sentence plausibility				
Intercept	335.6	1.40	239.12	<.001*
SemD	-0.03	0.33	-0.08	.937
Group Allocation	-0.60	1.29	-0.46	.645
SemD x Group	-0.46	0.33	-1.40	.162
(c) Semantic relatedness				
Intercept	313.08	1.54	203.44	<.001*
SemD	-0.30	0.44	-0.69	.503
Group Allocation	-1.33	1.27	-1.05	.299
SemD x Group	0.44	0.36	1.22	.228

Note. To improve interpretability, estimate and standard error were multiplied by 100,000 in speeded recognition and by 100 in both sentence plausibility and semantic relatedness.

Sentence plausibility (Day 2). Mean accuracy and mean RTs are summarized in the right column of Figure 6.4.

Accuracy There was a significant main effect of semantic diversity (z = 2.79, p = .005) such that high-diversity words (M = 71.5%) were judged more accurately than their low-diversity counterparts (M = 66.3%). There was also a significant main effect of group allocation (z = -3.43, p < .001) such that participants in the with-definition group (M = 72.5%) gave more accurate judgement than those in the without-definition group (M = 65.1%). The interaction between semantic diversity and group allocation, however, was not significant (z = 0.32, p = .748) (see Table 6.3 for model summary).

Reaction Time There were no effects of semantic diversity (t = -0.08, p = .937) or group allocation (t = -0.46, p = .645) and their interaction was also not statistically significant (t = -1.40, p = .162) (see Table 6.4 for model summary).

Semantic relatedness (Day 2). Mean accuracy and mean RTs in the two diversity conditions across the two groups are summarized in the right column of Figure 6.5.

Accuracy Neither semantic diversity (z = -0.96, p = .337) nor group allocation (z = -1.88, p = .060) had a significant effect on judgement accuracy, and their interaction was also not statistically significant (z = -1.05, p = .292) (see Table 6.3 for model summary).

Reaction Time Once again, there was no effect of semantic diversity (t = -0.69, p = 503) or group allocation (t = -1.05, p = .299) and no interaction (t = 1.22, p = .228) (see Table 6.4 for model summary).

2.2.3. Results summary

Day 1. Semantic diversity had an opposite effect on speeded recognition across the with- and without-definition groups. In the latter, novel words experienced in the low-diversity condition were recognized with greater accuracy, mirroring the findings from Experiment 1 (Chapter 2). In contrast, in the with-definition group, novel words experienced in the high-diversity condition were recognized faster than their low-diversity counterparts by approximately 17ms, replicating the 21-ms advantage reported in Experiment 2 (Chapter 3) and the 16-ms advantage in Experiment 3 (Chapter 5).

In sentence plausibility, participants in the with-definition group gave more accurate judgement to novel words experienced in the high-diversity condition, suggesting that the high- (vs. low-) diversity words were more context-independent. On the other hand, semantic diversity did not have an effect on plausibility judgement in the without-definition group.

Finally, in semantic relatedness, there was evidence that high-diversity words were disadvantaged, regardless of whether they were learned alongside a definition or not. In the with-definition group, accuracy was equally good across the diversity conditions but the high-diversity words elicited significantly slower response. In the without-definition group, accuracy rate was lower in the high- (vs. low-) diversity condition, although RTs were not different across conditions.

In summary, results from Day 1 are generally in line with the pre-registered hypotheses: (1) The provision of a familiar definition can serve to anchor, enabling high-diversity words to be better placed to take advantage of contextual variation. A result of this is faster identification and greater ability to fit into new contexts (i.e., greater context-

independence). (2) When a familiar definition is absent, novel words experienced in low-diversity contexts would enjoy a learning advantage because the repeated familiar topic helps reinforce and stabilize the novel words (i.e., anchoring).

Day 2. There was no detectable effect of semantic diversity in either speeded recognition or relatedness judgement. In plausibility judgement however, there was a clear effect of semantic diversity in both groups, such that high-diversity words received more accurate judgement, regardless of whether or not a familiar definition was present at exposure.

Results from Day 2 are more difficult to interpret, not only because two out of the three experimental tasks yielded nulls results, but also because the test phase on Day 1 afforded participants additional exposure to the novel words in isolation (via the speeded recognition test) and in neutral contexts (via the sentence plausibility test). It is impossible to know whether this impacted on the effect of semantic diversity induced at exposure and perhaps reducing any effect of semantic diversity on Day 2.

2.3. Discussion

In Experiment 4, I asked whether the provision of an explicit definition could help anchor novel words, thereby enabling them to be better placed to take advantage of contextual variation. One hundred young adults from Prolific were randomly allocated to learn novel words with or without a definition composed of a familiar concept (e.g., *ossanic* = not telling the truth). Regardless of which group they were allocated to, participants experienced the novel words in passages that spanned either four (high semantic diversity) or one topic (low semantic diversity). Frequency of exposure to novel words was identical across the diversity conditions. Word learning was assessed immediately after exposure and one day later, using three different experimental measures: speeded recognition, sentence plausibility, and semantic relatedness. Results from Day 1 are largely in line with the pre-registered hypothesis that the provision of a meaningful definition can serve to anchor. These will be considered first, beginning with the results from the *without-definition* group.

2.3.1. Day 1

Low-diversity advantage in the without-definition group

The experimental design in the without-definition condition is very similar to that of Experiment 1: Participants had to infer the meanings of novel words that were presented in

passages spanning one or multiple familiar topics (six in Experiment 1, four in Experiment 4). Contrary to Experiment 1 though, participants in Experiment 4 were explicitly instructed to learn the target words, which were presented alone before they were read in a passage. Experiment 4 also recruited participants from more heterogenous backgrounds. Despite these differences, the results of Experiments 1 and 4 converged to show a clear low-diversity advantage such that novel words experienced in the one-topic condition were recognized with greater accuracy in speeded recognition and were judged more accurately in semantic relatedness. These results, therefore, provided further support for the notion of anchoring, which argues that experiencing novel words repeatedly in a familiar topic enables the novel words to form strong and reliable associations with a pre-existing concept in long-term memory, thereby reinforcing and stabilizing the target's mental representations.

Restricting diversity in the initial stage of exposure (i.e., starting small) may facilitate the learning of linguistic knowledge, as seen in previous investigations of word learning (Horst et al., 2011; Jorge-Botana et al. 2017; Scott & Fisher, 2012), morphology (Tamminen et al., 2015) and syntax (Elman, 1993). In children, restricting diversity may be beneficial because it reduces cognitive load (Horst, 2013) and but in adults, restricting diversity may aid learning because it gives the new knowledge a chance to form secure associations with pre-existing concepts, a process proposed to facilitate long-term retention in adults (Mak, 2019). In sum, the without-definition group in Experiment 4 provided further support for the notion that in the early stages of word learning from reading, *less* diverse contexts can facilitate learning, perhaps by offering readers more opportunities to anchor the new words onto pre-existing knowledge.

Learning novel words alongside a familiar definition can serve to anchor

In the with-definition group, participants viewed a novel word alongside a meaningful definition (e.g., *not telling the truth*) before experiencing the word in a passage. Contrary to the low-diversity advantage seen in the without-definition group, participants in the with-definition group showed a high-diversity advantage in word recognition. This contrasting effect of semantic diversity suggests that the provision of a familiar definition helped anchor the novel words and enabled them to be better placed to take advantage of contextual variation.

As discussed in earlier chapters, there are multiple, and potentially related explanations as to why contextual variation might boost word recognition. The first is that words experienced in more diverse contexts are better encoded, resulting in more

accessible orthographic representations. A second account is that experiencing a word in more diverse contexts gives rise to a mental representation that is more context-independent (Jones et al., 2012). These accounts are not mutually exclusive and may work in tandem to bring about the high-diversity advantage seen in word recognition (Chapter 5). What remains unclear is how these accounts relate to each other; for instance, are they separate accounts? Or do they share the same underlying factor? This needs to be explored in future studies.

High-diversity words in the with-definition group can be integrated into new contexts more easily

Experiment 4 added the sentence plausibility judgement task, where participants read neutral sentences containing a newly learned word and judged whether or not the sentence made sense. The sentences, taken from Experiment 3, were contextually uninformative and did not reveal the meanings of the target words. In the with-definition group, high-diversity words were judged more accurately (M_{High} = 74.7% vs. M_{Low} = 67.0%), indicating that experiencing a word in diverse contexts during exposure better equipped readers to understand it in a neutral context (Joseph & Nation, 2018). This processing advantage fits nicely with the eye movement data reported in Experiment 3 and by Pagán and Nation (2018), both of which showed that newly learned high-diversity words are easier to process in neutral sentences. Taken together, these findings support the notion that words appearing in more diverse contexts tend to become more context-independent, enabling them to be processed with greater efficiency in never-encountered contexts. Note, however, that in the without-definition group, there was no reliable difference between the high- and low- diversity conditions in either accuracy or RTs. This suggests that diverse contexts alone may not be sufficient to bring about greater context-independence immediately in newly learned words.

Meanings of high-diversity words tended to be less precise and less accessible

The Semantic Distinctiveness Model (Jones et al., 2012) predicts that words appearing in more diverse contexts tend to have less stable semantic representations. This would make them more difficult to judge in tasks requiring a specific semantic detail (Chapman, 2019; Hoffman & Woollams, 2015; Hsiao et al., 2020; Johns et al., 2016). Data from the semantic relatedness task in Experiment 4 are generally in line with this prediction.

In the without-definition group, accuracy rate was significantly lower in the high-diversity condition (M_{High} = 67.1% vs. M_{Low} = 72.4%), replicating the findings from Experiment 1 and Johns et al. (2016). This disadvantage, however, was not present in the with-definition

group (M_{High} = 75.3% vs. M_{Low} = 75.5%). This suggests that the provision of an explicit definition, perhaps unsurprisingly, boosted the semantic precision of high-diversity words (Bolger et al., 2008). An implication of this is that the provision of definitions may facilitate the learning of the meanings of high-diversity words more than of low-diversity words, potentially explaining why the effect of glosses on word learning is not always consistent across studies (e.g., Jacobs et al., 1994; Watanabe, 1997; Wilkinson & Houston-Price, 2013). However, although the high-diversity words were not disadvantaged in terms of accuracy in the with-definition group, they elicited significantly slower judgement, indicating poorer accessibility (M_{Low} = 1528 ms vs. M_{High} = 1614 ms). This suggests that while providing explicit definitional knowledge has value and can enhance semantic precision, it has limits (Bolger et al., 2008) and should not be considered a silver bullet to word learning (Chapter 1; Section 2.2.2). In sum, data from semantic relatedness are largely in line with previous studies in showing that high-diversity words are more difficult to judge in this kind of semantic tasks. This might be because high-diversity words are more nuanced in meaning and/or less well-integrated into prior knowledge.

2.3.2. Day 2

It is difficult to make sense of the data from the delayed session, not least because two of the three tasks yielded null results. Furthermore, the additional exposure afforded by the test phase on Day 1 might have contaminated the contextual history of the target words and reduced any effect of semantic diversity in the delayed session. Therefore, the interpretations in this section are, at best, speculative in nature.

Semantic diversity and offline consolidation

In the without-definition group, semantic diversity had no effect on plausibility judgement on Day 1 (M_{High} = 64.5% vs M_{Low} = 64.9%). However, on Day 2, accuracy rate in the high-diversity condition increased to 68.0% while that in the low-diversity condition decreased to 62.3%. The difference between the diversity conditions was statistically significant (z = -2.18, p = .029). The emergence of a high-diversity advantage on Day 2, therefore, hints at the possibility that semantic diversity may interact with offline consolidation. In this section, I focus on the *without-definition* group.

It is well-documented that a period of sleep-related offline consolidation (vs. wakefulness) can strengthen the memory representations of newly learned words (e.g., Henderson, Devine, Weighall, & Gaskell, 2015; James et al., 2019) and facilitate their

integration into existing knowledge (Bowers, Davis, & Hanley, 2005; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003; Wang et al., 2017). If offline consolidation strengthens the memory representations of newly learned words, why was improvement seen only among the high-diversity words in plausibility judgement? Assuming this is not a spurious finding, I suggest that high-diversity words might be preferentially strengthened during offline consolidation. Adelman et al. (2007) and Jones et al. (2012) argued that words appearing in more diverse contexts are more likely to be needed in future unknown contexts, so they should be more relevant and important for the future than words that only appear in a restricted set of contexts. Three studies investigating the effect of sleep-related offline consolidation (van Dongen, Thielen, Takashima, Barth, Fernández, 2012; Wamsley, Hamilton, Graveline, Manceor, & Parr, 2016; Wilhelm et al., 2011) demonstrated that newly learned materials that are more relevant for the future are prioritized for consolidation. This suggests the possibility that high-diversity words might have been selectively enhanced during offline consolidation. If high-diversity words were indeed preferentially enhanced, this would explain why they were advantaged in the plausibility task on Day 2 but not on Day 1. In fact, there was also a trend for high-diversity words to be more accurately recognized in speeded recognition on Day 2 (M_{High} = 95.8% vs. M_{Low} = 91.1%). Although this difference did not reach statistical significance (z = 0.93, p = .354), it stands in contrast to the high-diversity *disadvantage* in recognition accuracy on Day 1 (M_{High} = 93.5% vs. M_{Low} = 95.9%; z = 1.97, p = .0495). This discrepancy hints at a potential interaction between semantic diversity and offline consolidation. To sum up, the plausibility data from the without-definition group open up the exciting possibility that offline consolidation may preferentially enhance lexical representations that were encountered in more diverse contexts. Note that the high-diversity words in the without-definition group were not anchored to a definition during exposure on Day 1, so the fact that a high-diversity advantage emerged in plausibility judgement on Day 2 negates the possibility that anchoring is a pre-condition to a high-diversity advantage.

Although there is little doubt that offline consolidation influences memory, an alternative explanation is possible. Relative to low-diversity words, words appearing in more diverse contexts in natural language tend to be more polysemous and semantically nuanced (Hoffman & Woollams, 2015; Hsiao et al., 2020; Johns et al., 2016). Extending this to the word learning situation, nuance in meaning might have induced more processing effort on the high-diversity words in the plausibility and relatedness judgement tasks on Day 1. This might have strengthened the semantic representations of the high-diversity words by the

end of the Day 1 test phase. A potential consequence of this would be a greater performance boost in the Day 2 test phase. Currently, it is not possible to rule out this possibility as all participants completed the test phase on Day 1. Future studies should consider eliminating the test phase on Day 1 to give a clearer picture on how semantic diversity may interact with offline consolidation.

The effect of semantic diversity in the with-definition group on Day 2

Participants who had the opportunity to map the novel words to a meaningful definition showed a high-diversity advantage in recognition speed on Day 1. However, this advantage was not evident on Day 2 (M_{High} = 624 ms vs. M_{Low} = 628 ms). It is unclear why this was the case. One plausible explanation is reduced statistical power: only 40 of the 50 participants in the with-definition group took part in the delayed session. Another plausible explanation is that both the high- and low-diversity words were encountered in isolation (in speeded recognition) and in uninformative contexts (in plausibility judgement) in the Day 1 test phase. This might have contaminated the contextual history of the target words, which might, in turn, neutralize the effect of semantic diversity induced at exposure. However, this explanation is at odds with the high-diversity advantage in plausibility judgement, which was present on both days in the with-definition group. It is unclear why one advantage disappeared and another remained across days. Potentially, speeded recognition tapped the learning of form while plausibility judgement tapped the learning of meaning, and speculatively, the latter might be more resistant to the neutralizing effect of the Day 1 test phase. To complicate the picture further, this proposal actually contradicts the result from relatedness judgement: If the learning of meaning was more resistant to the neutralizing effect of the Day 1 test phase, the high-diversity disadvantage in relatedness judgement should be evident on both Days 1 and 2. However, the high-diversity words were judged equally well as their low-diversity counterparts on Day 2. Presently, it is difficult to make sense of these inconsistencies. Future studies are needed to examine the longer-term effect of semantic diversity, perhaps by having “Day” as a between-participant manipulation.

Chapter 7

General Discussion

This thesis was inspired by the finding that item-level contextual history, as captured by corpus-derived statistics such as semantic diversity, influences lexical processing: Known words that have been experienced in more varied linguistic contexts are easier to recognize in tasks such as lexical decision but more difficult to reflect on in tasks such as semantic judgement (e.g., Hoffman & Woollams, 2015; Hsiao et al., 2020; Jones et al., 2012). While consistent in showing an effect of contextual variation, research using corpus-derived metrics cannot explain the nature of its effect, nor how it develops and accrues over time, as words are experienced. Instead, word-learning experiments are needed, where researchers can manipulate contextual experience directly and track its effects on learning. I adopted this approach in four behavioral experiments, complemented by three computational simulations. In the behavioral experiments, young adults experienced novel pseudowords (e.g., *ossanic*) that replaced English words (e.g., *mendacious*) in either high or low diversity passages: High-diversity passages spanned multiple distinct topics while low-diversity passages centered on one discourse topic. With frequency (and document count) held constant and the target words counterbalanced across conditions, it was possible to measure how contextual history, as indexed by semantic diversity, influenced word learning.

1. Theoretical contributions

1.1. Anchoring and the learning of word form

In Experiment 1, pseudowords experienced in the low-diversity passages outperformed their high-diversity counterparts in the two tasks that tapped the learning of form (i.e., pseudoword inferiority & speeded recognition). I attributed this finding to anchoring, the idea that experiencing novel words repeatedly in a familiar topic can secure the novel words onto the topic's representation in long-term memory. In turn, the pre-existing representation helps reinforce and stabilize the novel words. Results from Simulation 1 provided computational support for the utility of anchoring in the initial stages of word learning and suggested that anchoring could be conceptualized as higher cluster coefficients (i.e., more close-knit immediate neighborhoods) which can facilitate activations to circulate back to the target word.

Simulation 2 and Experiment 2 showed that a high-diversity advantage in word recognition, as shown in experiments using corpus-derived metrics (e.g., Hsiao & Nation,

2018; Jones et al., 2012), may emerge if novel words were given an anchoring opportunity before contextual variation was introduced. Specifically, I argued that an anchoring opportunity allowed the target words to become better placed to take advantage of contextual variation. However, given the results from Johns et al. (2016) (and Experiment 4 reported in Chapter 6), anchoring should not be interpreted as a pre-condition to the high-diversity advantage in recognition. Instead, it should be seen as a facilitator or a contributor. This recognition advantage among high-diversity words was then replicated in Experiment 3, where the eye-tracking data showed that these words might have been encoded more strongly during exposure and gained greater context-independence.

In Experiments 1 to 3, novel words were all anchored to a familiar topic. Experiment 4 showed that an explicit definition composed of familiar concepts (e.g., ossanic = *not telling the truth*) could also serve to anchor. In line with the results from Experiments 2 and 3, novel words initially anchored to a familiar definition were better able to take advantage of contextual variation, allowing them to be recognized more efficiently than counterparts that were also anchored to a definition but remained in one topic throughout exposure. Summing across all four behavioral experiments, my findings suggest that both familiar topics and definitions can serve to anchor novel words onto pre-existing knowledge.

An important question that needs to be addressed is how anchoring fits with existing models of lexical processing, especially the Semantic Distinctiveness Model (Jones et al., 2012; 2017). According to this model, when a word is experienced, the encounter will be compared to its contextual history stored in memory; if the current context is distinct from the word's contextual history, there is more pressure for stronger encoding. One crucial element in this model is the comparison between current and stored contexts. In the case of newly learned words, I argue that such a comparison may not take place because the contextual history of newly learned words may not be well-established or easily retrievable enough to enable comparison. With some anchorage in place however, the contextual history may be stabilized and become more retrievable, making comparison between current and stored contexts more likely. This might then trigger stronger encoding when the current context is distinct, enabling the words to become more context-independent. In turn, a high-diversity advantage in word recognition may emerge. Therefore, I posit that the notion of anchoring does not contradict the Semantic Distinctiveness Model; instead, the Model needs to be refined and expanded such that it can accommodate newly learned words that are likely to have weak and impoverished contextual history.

Finally, I consider the extent to which anchoring is psychologically plausible. Specifically, is there any evidence showing that high-diversity words are initially experienced in a restricted set of contexts? Given the reading phase in my experiments took place in just one session, I am cautious not to make any bold claims with regard to how and when anchoring might happen in real life. However, it is possible to note some relevant observations. For instance, young children frequently ask their caregivers to repeat a storybook, or replay a video (Crawley, Anderson, Wilder, Williams, & Santomero, 1999; Sulzby, 1985; Yanek, 2018), and indeed, word learning from a story is enhanced when toddlers hear the same story verbatim (Horst et al., 2011; Williams & Horst, 2014; see also Childers, Paik, Flores, Lai, & Dolan, 2017; Scott & Fisher, 2012). Second, and more generally, language experience takes place in more restricted contexts early in development with diversity growing as children engage with more speakers and in different communicative contexts (see for example, Roy, Frank, DeCamp, Miller, & Roy, 2015); learning to read is also likely to bring variation and diversity to a word's contextual history (e.g., Dawson et al., in press; Montag, Jones, & Smith, 2015; Nation, 2017). This developmental trend and the notion of anchoring converge in suggesting that initially restricting a word's diversity may facilitate subsequent learning (Jorge-Botana et al., 2017; see also Elman, 1993). Clearly however, future experimental and computational work is needed to capture lexical and contextual history and to evaluate the extent to which developmental trends overlap with my notion of anchoring.

1.2. Contextual variation

A large body of empirical evidence suggests that in addition to frequency, the context in which a word has been experienced influences verbal learning and memory (e.g., Hoffman & Woollams, 2015; Hsiao et al., 2020; Jones et al., 2017; Nation, 2017). As noted in the beginning of Chapter 2, it is not clear whether it is necessary to take the semantic content into account when conceptualizing contextual variation (Jones et al., 2012). If semantic content did not matter, there should be no learning difference between the high- and low-diversity conditions in the current experiments (and simulations) as target words in the two conditions were matched on both frequency and document count. Yet, the two diversity conditions clearly differed in how they influenced learning across experiments. Therefore, the findings reported in this thesis provide a strong case for taking the semantic content into account when conceptualizing and computing contextual variation (e.g., Jones et al., 2012; Hollis, 2020; Hsiao & Nation, 2018).

Contextual variation has been argued as a key to achieving high-quality lexical representations (Bolger et al., 2008; Nation, 2017; Perfetti, 2007). However, at present, evidence from word learning experiments does not paint a coherent picture. Some studies found diverse contexts to be more beneficial to word learning (e.g., Bolger et al., 2008; Frances et al., 2020; Johns et al., 2016; Rosa et al., 2017); others reported mixed results (e.g., Wilkinson & Houston-Price, 2013), no effect (e.g., Joseph & Nation, 2018), or a negative effect (e.g., Chilton & Ehri, 2015; Horst et al., 2011; Tamura, 2015; Williams & Horst, 2014). This inconsistency is not surprising as there are various factors that may have influenced any effect of contextual variation across these studies, including, but not limited to, age of participants (toddlers vs. undergraduates), word class of the target words (noun vs. verb), training paradigm (intentional vs. incidental), duration of training, and whether the non-diverse condition was implemented as verbatim repetition (Bolger et al., 2008; Pagán & Nation, 2019). One particularly crucial factor, as demonstrated by Experiment 1 in this thesis is topic familiarity. My experiment used more familiar topics while Johns et al. (2016) used more obscure topics, and the effect of semantic diversity (on the learning of word form) was opposite across experiments. This highlights the need for experimenters to take participants' prior knowledge into consideration when designing the study (Mak, 2019), as this often has a substantial effect on the learning outcome (see for example, Greve et al., 2019; Havas et al., 2018; Recht & Leslie, 1988). Future studies are needed to directly examine the interaction between semantic diversity and topic familiarity, perhaps using a 2 x 2 within-subject design.

Another factor that may modulate the effect of semantic diversity on lexical processing is how well-established a word's mental representation is. It is generally accepted that word knowledge exists on a continuum—from very weak and impoverished (e.g., a newly learned word) to very accessible and complex (Horst, 2013), and each encounter with a word—be it in spoken discourse or in text—offers a learning opportunity to enrich and refine its representations (Nation, 2017). Diversity in context may not be beneficial in the earliest stage of word learning, because the weak representations that characterize newly encountered words may benefit more from reinforcement and stabilization via more similar (and topically familiar) contexts rather than enrichment via diverse contexts. In contrast, as showed in Experiments 2 to 3, given an anchoring opportunity, further enrichment via semantically diverse contexts can enhance learning, even in the early stages of learning a new word. This strengthens the idea that the effect of semantic diversity on word learning hinges upon a third factor, providing a plausible explanation as to why there is inconsistency

in the current "word learning x contextual diversity" literature. One serious gap in the evidence base is the lack of data from longer-term experiments that manipulate contextual history directly and monitor its effects on learning and processing. These are very much needed to inform our understanding of precisely what a corpus-derived metrics such as semantic diversity reflect.

Finally, one factor that merits a brief discussion is the length of the text in the reading phase. This varied substantially across the four experiments presented in this thesis: Experiment 1 used longer passages (*M* = 127.0 words); Experiment 2 used a mixture of short (*M* = 36.2 words) and long passages (*M* = 125.8 words), and finally, Experiments 3 and 4 made use of short passages only (*M* ≈ 35.8 words). Despite this difference, the effect of anchoring and contextual variation remained relatively consistent across experiments. For instance, the benefits associated with anchoring were observed in both Experiment 1 and the without-definition group in Experiment 4, and Experiment 3 replicated the high-diversity advantage in Experiment 2. Together, they highlight the robustness of the current findings and the fact that variation even at the level of sentences influences learning.

1.3. Learning of meaning

In line with the existing literature (e.g., Chapman, 2019; Hoffman & Woollams, 2015; Hsiao et al., 2020; Johns et al., 2016; Jones et al., 2012), Experiment 1 (Chapter 2) and Experiment 4 (Chapter 6) showed that the meanings of novel words appearing in more diverse contexts were more difficult to judge in the semantic relatedness judgement task. This disadvantage is usually interpreted as indicating that the meanings of high-diversity words are flexible enough to fit various contexts (Bolger et al., 2008), making them more polysemous and semantically nuanced. A consequence of this is that they become more difficult to judge in tasks requiring specific semantic detail. While this interpretation has both behavioral and computational support (Johns et al., 2016), an alternative explanation is also possible: I proposed that when matched on frequency, a high-diversity word had fewer opportunities to be mapped onto each of its discourse topic than a low-diversity word; a consequence of this may be poorer integration into pre-existing semantic knowledge. In turn, this may make semantic judgement on high-diversity words more challenging. As shown in Simulation 3a, high-diversity words were consistently more distant from their surrounding words in semantic space than their low-diversity counterparts. This suggests that the high-diversity words may be less well integrated into the discourse topics. Currently, it is difficult to tease apart the semantic complexity account and my integration account, because it was not

possible to match a high-diversity and a low-diversity word on both frequency and the number of times they are experienced in each topic. To evaluate these two accounts, future studies may consider matching high-diversity and low-diversity words on not only frequency but also the number of senses.

Finally, it is important to note that in Experiments 2 (Chapter 3) and 3 (Chapter 5), high-diversity words were judged as efficiently as their low-diversity counterparts in semantic relatedness. This may be because these high-diversity words were experienced in only two discourse topics, as opposed to six topics in Experiment 1, thereby making them less nuanced and enabling them to be better integrated into pre-existing knowledge. A potential real-life implication for explicit teaching is that the nuances of high-diversity words might best be introduced to language learners gradually, instead of all at once.

2. Methodological contributions

Findings from the behavioral experiments reported in this thesis show that it is possible to manipulate context and induce variation in word learning in a single laboratory session. Clearly, such learning is some distance from the complex item-level knowledge about a word that builds slowly over time via natural language experience situated in the world. Nevertheless, the advantage of the current approach is that it allowed the learning environment to be carefully controlled and the effects of semantic diversity to be directly measured.

2.1. Pseudoword inferiority as an index of word learning

Experiments 1 and 2 introduced the pseudoword inferiority effect (Grainger & Jacobs, 2005) as an index of word learning. Building from the premise that words generate a top-down influence that can interfere with letter identification (Reicher, 1969; Wheeler, 1970), if a newly learned word has become well-established, it should induce a pseudoword inferiority effect. Evidence for this was seen in both Experiments 1 and 2, showing that six to eight exposure to a new form in incidental reading is sufficient to drive changes in performance on the pseudoword inferiority task. The size and direction of the effect was also sensitive to the contextual manipulation: In Experiment 1, novel words encountered in the low-diversity condition showed more interference than those in the high-diversity condition. In Experiment 2, the direction of the effect flipped, with high-diversity words bringing about more interference. These findings mirror those seen in the speeded recognition task in both experiments, attesting to the validity of pseudoword interference as an index of word

learning. Finally, although accuracy in the novel word condition (e.g., OSSAMIC) of the pseudoword inferiority tasks did not correlate with accuracy in speeded recognition (Experiments 1 and 2: *r*s ≈ -.20, *p*s > .05)[15], it correlated moderately with that in semantic relatedness judgement (Experiment 1: $r = -.40$, $p = .011$; Experiment 2: $r = -.64$, $p = < .001$).[16] This provides further evidence for pseudoword interference being a valid index of word learning.

2.2. Computational simulations

The network simulations reported in Simulations 1 and 2 demonstrated the utility of the spreadr package (Siew, 2019) to simulate spreading activation as a possible mechanism to explain the effects of anchoring and contextual variation. Computational simulations offer an opportunity to replicate behavioral findings and to study the independent variable (e.g., semantic diversity) without concern over factors that are difficult to control for in a behavioral study (e.g., arousal, level of background knowledge, pre-existing attitude regarding characters or situations). In addition, they can also inform experimental design. For instance, there were numerous potential ways to implement semantic diversity in Experiment 2 (e.g., number of topics, number of times a word is experienced in each topic). Using spreadr helped inform the design of the experiment by narrowing down these possibilities. Of course, my simulations have their limitations. For instance, Simulations 1 and 2 were built upon assumptions derived from spreading activation dynamics, where each word is conceptualized as a single node (Collins & Loftus, 1975), and other possible accounts based on other assumptions are also plausible (e.g., words as distributed representations; e.g., Plunkett, 1998).

Simulation 3 presented a mathematical approach to examine the possibility that high-diversity words are less well-integrated into prior semantic knowledge, thereby making them more challenging to judge in relatedness judgement. In this simulation, I transformed the words in the reading materials from Experiments 1 and 2 into word vectors and mapped them onto semantic space, enabling the quantification of how close the target words are to their neighboring words. I inferred that the closer the neighboring words are to the target words, the greater the likelihood of the target words being integrated into prior knowledge.

---

[15] This may be because accuracy in speeded recognition was heading towards ceiling.

[16] Note that poorer accuracy in the pseudoword inferiority task indicates greater top-down influence, and hence, *better* learning of the novel words. It is therefore not surprising that accuracies in pseudoword inferiority and semantic relatedness are *negatively* correlated.

Simulations 3a and 3b, using the same procedure, produced results that mirrored those from human learners in Experiments 1 and 2. This is encouraging and shows promise that this novel approach could be used to predict semantic learning. More empirical studies are needed to evaluate the utility of this novel approach in predicting semantic learning and semantic integration.

2.3. Experiment 4 and online testing

Experiment 4 recruited participants online via Prolific owing to COVID-19 restrictions. Results from this experiment were generally in line with those from the lab-based studies (Experiments 1 to 3). For instance, in the without-definition group, there was a clear low-diversity advantage in recognition accuracy on Day 1, replicating the results from Experiment 1 and highlighting the utility of anchoring in the early stage of word learning. One major strength of Experiment 4 was that participants were recruited from more heterogeneous backgrounds (i.e., not restricted to convenience samples of Oxford psychology undergraduates). Therefore, the fact that Experiment 4 found comparable results to those from Experiments 1 to 3 gives some confidence that the current findings are robust and generalizable.

However, there are a few commonly held concerns regarding online studies (Hulme, 2018). For example, how can data quality be guaranteed when the participants were not monitored? I addressed this in Experiment 4 by using regular attention checks and removing participants who failed a percentage of those checks, as set out in the pre-registered study protocol. In addition, the accuracy data collected for Experiment 4 were normally distributed, and the absence of a ceiling effect offers some reassurance that participants didn't cheat (e.g., by jotting the target words down). Rodd (2019) also offers reassurance that online data collection has fidelity, if experiments are constructed carefully.

Another concern about online studies is the reliability of the RT data. Yet several empirical studies have now demonstrated that RT data from online platforms (e.g., Gorilla) are highly comparable to those from lab-based studies (e.g., Anwyl-Irvine, Dalmaijer, Hodges, Evershed, 2020; Barnhoorn, Haasnoot, Bocanegra, & van Steenbergen, 2015). This was also the case in my experiments with the RT advantage in speeded recognition being similar in magnitude between Experiment 3 (lab-based; 16-ms) and 4 (online; 17-ms).

3. Limitations and future directions

In this section, I outline some critical limitations in the behavioral experiments and highlight some areas that warrant further investigation.

First, across all four experiments, participants experienced the target words in a single exposure phase that took no longer than 45 minutes. Clearly, this is some distance from reality, where word knowledge accrues over time and over repeated exposure. Future longitudinal work is needed to follow up on how contextual variation influences the learning and accumulation of word knowledge. This will also help strengthen the understanding of precisely what corpus-derived metrics such as semantic diversity reflect (e.g., Cevoli et al., 2020; Hollis, 2020; Hsiao & Nation, 2018). Experiment 4 attempted to investigate the longer-term effect of semantic diversity in a delayed session that took place 24 hours after the exposure phase. The evidence, however, was weak and inconclusive. This might be because repeated testing had neutralized the effect of semantic diversity induced at exposure. Future studies interested in the longer-term effect of semantic diversity should consider having "Day" as a between-subject manipulation or testing only a portion of the word targets in each session. In addition to longitudinal studies, learning experiments with good ecological validity are needed to shed light on how the current findings can be translated into the real world. For instance, the results from Experiments 1 to 3 suggest that children and second language learners would better acquire the meanings of high-diversity words if they were initially encountered in a restricted set of topics (as opposed to experiencing them in multiple topics). This may be because restricting diversity helps reduce the words' semantic complexity and provide more opportunity for the emerging words to be integrated into long-term memory. It would be of great value if this prediction can be tested in a language classroom with carefully selected reading materials and with real, instead made-up, words. As part of my ongoing postdoctoral project, I am running a large-scale online comprehension program with English-speaking children, with the aim of testing whether the results from this thesis can be translated into more naturalistic settings.

Relevant to the last point about using real instead of fake words, the experiments reported in this thesis followed Johns et al. (2016) in using a mixture of nouns, verbs, and adjectives as the base words (e.g., *mendacious*, *extirpate, constellation*). Although this is not a limitation *per se*, it would be important for future studies to systematically investigate whether part of speech or concreteness of the target words mediate the effect of semantic diversity. For instance, all the target words in Joseph and Nation (2018) were abstract verbs (e.g., *exacerbate*, *languish*) but no effect of semantic diversity was observed in their study;

in contrast, Rosa et al. (2017) made use only of concrete nouns (e.g., *canines* in Spanish) and showed a high-diversity advantage in both the learning of form and meaning. These contrasting results may be related to the degree of abstractness associated with verbs and nouns. Future studies are needed to directly examine how semantic diversity may interact with the abstractness of the target words.

Consider Joseph and Nation (2018) once more. Despite sharing a similar experimental design as Experiment 1, their study found no evidence that learning differed between their high- and low-diversity conditions. This stands in contrast to the findings from Experiment 1, which showed that low-diversity contexts may facilitate initial learning via anchoring. The contrasting results may be related to the degree of semantic overlap among the low-diversity passages: In Experiment 1, the topics of the low-diversity passages were mostly restricted to one hyponym; that is, if a set of low-diversity passages were about Kate Middleton, they were only related to Kate Middleton, not other members of the royal family. In contrast, in Joseph and Nation's (2018) experiment, the low-diversity passages spanned topics about various hyponyms (e.g., horse, canary, dog). Future work is needed to investigate how much semantic overlap is necessary in the low-diversity condition to bring about anchoring, and whether the effect of semantic overlap may be different across children and adults, given the schemata (i.e., knowledge structures) in children are still developing.

In addition to children vs. adults, future studies may also consider looking into how semantic diversity may interact with individual differences in for example verbal working memory. Recall that Simulation 3b found no consistent effect of semantic diversity across window sizes on the relatedness values. Specifically, when the window size was small (= 8), target words in the low-diversity condition had higher relatedness values (suggestive of better semantic integration), but when the window size increased (= 16), the effect of semantic diversity flipped such that high-diversity words had higher relatedness values. Speculatively, if the window size can be considered as a human-equivalent of verbal working memory, Simulation 3b suggests that the effect of semantic diversity may interact with working memory capacity to produce different learning outcomes. Future studies are needed to examine this possibility.

Finally, two key explanations were put forth to account for why high semantic diversity might be associated with more efficient word recognition: the encoding and the context-independence accounts. The eye-tracking data from Experiment 3 (Chapter 5) gave weight to both. However, the relationship between these accounts remains to be fully

explored. It is entirely possible that they are separate accounts, or one causes the other, or they stem from the same underlying factor. Longer-term learning experiments and computational modelling would help elucidate the relationship between these accounts.

4. Conclusion

From mid-childhood onwards, new words are primarily acquired via independent reading. Numerous studies have now demonstrated that the ease with which a word is learned during reading is influenced by the characteristics of the context in which the word occurs. This thesis examined the effect of one of these characteristics: contextual variation, as indexed by semantic diversity. As demonstrated in four behavioral experiments with young adults and three computational studies, it is evident that the effect of semantic diversity on word learning is not clear-cut and is oftentimes affected by other factors such as topic familiarity and the provision of definitions. Overall, these studies demonstrated that (i) restricting the semantic diversity of a new word and capitalizing on learners' prior knowledge can facilitate word learning in the initial stages, potentially via anchoring, and (ii) novel words that have been given an opportunity to anchor to a pre-existing concept—operationalized as either a discourse topic or an explicit definition—are better placed to take advantage of more semantically diverse contexts.

References

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity not frequency determines word naming and lexical decision times. *Psychological Science*, *17*(9), 814–823.

Altszyler, E., Ribeiro, S., Sigman, M., & Fernández Slezak, D. (2017). The interpretation of dream meaning: Resolving ambiguity using Latent Semantic Analysis in a small corpus of text. *Consciousness and Cognition*, *56*, 178–187.

Anderson, J., Anderson, A., Lynch, J., & Shapiro, J. (2004).Examining the effects of gender and genre on interactions in shared book reading. *Literacy Research and Instruction, 43*, 1–20.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*, 261–295.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*, 703–719.

Anderson, J. R., & Schooler, L.J. (1991). Reflections of the environment in memory. *Psychological Science, 2*, 396–408.

Anderson, R., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (Ed.), Comprehension and teaching: *Research reviews* (pp. 77–117). Newark, DE: International Reading Association.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N. & Evershed, J. K. (2020) Gorilla in our midst: An online behavioral experiment builder. *Behavioral Research Methods, 52* , 388–407.

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.

Baglama, J., Reichel, L., & Lewis, B.W. (2019). irlba: Fast Truncated Singular Value Decomposition and Principal Components Analysis for Large Dense and Sparse Matrices. R package version 2.3.3. https://CRAN.R-project.org/package=irlba

Baker, S. K., Simmons, D. C., & Kame'enui, E. J. (1998). Vocabulary acquisition: Research bases. In D. C. Simmons, & E. J. Kame'enui (Eds.), *What reading research tells us about children with diverse learning needs* (pp. 183–218). Mahwah, NJ: Erlbaum.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283–316.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English lexicon project. *Behavioral Research Methods*, *39*(3), 445–459.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*(3), 438–481.

Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology, 30*, 1174–1220.

Barnhoorn, J. S., Haasnoot, E., Bocanegra, B. R., & van Steenbergen, H. (2015). QRTEngine: An easy solution for running online reaction time experiments using Qualtrics. *Behavior Research Method*, *47,* 918–929.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bates D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beck, I. L., McKeown, M. G., & McCaslin, E. S. (1983). Vocabulary: All contexts are not created equal. *Elementary School Journal*, *83*, 177–181.

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLoS ONE*, *6*(5), e19348.

Bergman Deitcher, D., Johnson, H., & Aram, D. (2019). Does book genre matter? Boys' and girls' word learning from narrative and informational books in the preschool years. *Journal of Research in Reading*, *42*(1), 193–211.

Biemiller, A., & Boote, C. (2006). An effective method for building vocabulary in primary grades. *Journal of Educational Psychology*, *98*, 44–62.

Blythe, H. I., Liversedge, S. P., Joseph, H. S., White, S. J., & Rayner, K. (2009). Visual information capture during fixations in reading for children and adults. *Vision Research*, *49*(12), 1583–1591.

Boers, F., Warren, P., He, L., & Deconinck, J. (2017). Does adding pictures to glosses enhance vocabulary uptake from reading? *System*, *66*, 113–129.

Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes*, *45*(2), 122–159.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, *24*(3), 127–135.

Bouchet-Valat, M. (2019). SnowballC: Snowball Stemmers Based on the C 'libstemmer' UTF-8 Library. R package version 0.6.0. https://CRAN.R-project.org/package=SnowballC

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbors: The impact of novel word learning on the identification of visually similar words. *Cognition*, *97*(3), 45–54.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science*, *27*(1), 45–50.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavioral Research Methods*, *41*(4), 977–990.

Carey, S. (1978). The child as word learner. In M. Halle, G. Miller, & J. Bresnan (Eds.), *Linguistic theory and psychological reality* (pp. 264e293). Cambridge, MA: MIT Press.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest*, *19*(1), 5–51.

Cervetti, G. N., Wright, T. S., & Hwang, H. J. (2016). Conceptual coherence, comprehension, and vocabulary acquisition: A knowledge effect? *Reading and Writing*, *29*(4), 761–779.

Cevoli, B., Watkins, C., & Rastle, K. (2020). What is semantic diversity and why does it facilitate visual word recognition? *Behavioral Research Methods*, *53,* 247–263.

Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning New Word Meanings from Context: A Study of Eye Movements. *Journal of Experimental Psychology: Learning Memory and Cognition*, *27*(1), 225–235.

Chan, K. Y., & Vitevitch, M. S. (2009). The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1934–1949.

Chan, K. Y., & Vitevitch, M. S. (2010). Network structure influences speech production. *Cognitive Science*, *34*, 685–697.

Chapman, C. A. (2019). How do word frequency and semantic diversity affect selection of representations in word processing? [Unpublished doctoral dissertation]. Rice University.

Chiesi, H. I., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge*. Journal of Verbal Learning and Verbal Behavioral*, *18*, 275–290.

Childers, J. B., Paik, J. H., Flores, M., Lai, G., & Dolan, M. (2017). Does Variability Across Events Affect Verb Learning in English, Mandarin, and Korean? *Cognitive Science*, *41*, 808–830.

Chilton, M. W., & Ehri, L. C. (2015). Vocabulary learning: Sentence contexts linked by events in scenarios facilitate third graders' memory for verb meanings. *Reading Research Quarterly*, *50*(4), 439–458.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80*, 183–198.

Church, K., & Gale, W. (1995) Inverse Document Frequency (IDF): A Measure of Deviation from Poisson. *Proceedings of the Third Workshop on Very Large Corpora*, 121–130.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual-route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Crawley, A. M., Anderson, D. R., Wilder, A., Williams, M., & Santomero, A. (1999). Effects of repeated exposures to a single episode of the television pro- gram Blue's Clues on the viewing behaviorals and comprehension of preschool children. *Journal of Educational Psychology*, *91*(4), 630–637.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695. http://igraph.org

Dawson, N., Hsiao, Y., Tan, A., Banerji, N., & Nation, K. (in press). Features of lexical richness in children's books: Comparisons with child-directed speech. *Language Development Research.* https://lps.library.cmu.edu/LDR/article/id/77/

Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., and Harshman, R. A. (1990). Indexing by latent semantics analysis. *JASIS*, *41*(6), 391–407.

Diakidoy, I. A. N. (1998). The role of reading comprehension in word meaning acquisition during reading. *European Journal of Psychology of Education*, *13*(2), 131–154.

Dumay, N., & Gaskell, M. C. G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychological Science*, *18*(1), 35–39.

Durso, F.T. & Shore, W.J. (1991). Partial knowledge of word meanings. *Journal of Experimental Psychology: General*, *120*, 190–202.

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, *48*(1), 71–99.

Floyd, S., & Goldberg, A. E. (2020). Children Make Use of Relationships Across Meanings in Word Learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *2*(999).

Firth, J. R. (1957). A synopsis of linguistic theory 1930–1955. In *Studies in Linguistic Analysis*. Philological Society. Reprinted in Palmer, F. (ed.) 1968. Selected Papers of J. R. Firth. Longman, Harlow.

Fischer, U. (1994). Learning words from context and dictionaries: An experimental comparison. *Applied Psycholinguistics, 15*, 551-574.

Frances, C., Martin, C. D., & Duñabeitia, J. A. (2020). The effects of contextual diversity on incidental vocabulary learning in the native and a foreign language. *Scientific Reports*, *10*(1), 13967.

Freebody, P., & Anderson, R. (1983). Effects on text comprehension of different proportions and locations of difficult vocabulary. *Journal of Reading Behavioral*, *15*, 19-29.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132.

Gaskell, M. G., Cairney, S. A., & Rodd, J. M. (2019). Contextual priming of word meanings is stabilized over sleep*. Cognition*, *182*, 109–126.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., … Yoon, H. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, *21*(3), 1–22.

Goldstein, R., & Vitevitch, M. S. (2014). The influence of clustering coefficient on word-learning: How groups of similar sounding words facilitate acquisition. *Frontiers in Psychology*, *5*(NOV), 2009–2014.

Grainger, J., & Jacobs, A. M. (2005). Pseudoword context effects on letter perception: The role of word misperception. *European Journal of Cognitive Psychology*, *17*(3), 289–318.

Greve, A., Cooper, E., Tibon, R., & Henson, R. N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. *Journal of Experimental Psychology: General*, *148*(2), 325–341.

Haastrup, K. (1991). *Lexical inferencing procedures or talking about words*. Tübingen, Germany: Gunter Narr.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662–720.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday lives of young American children*. Baltimore: Paul H. Brookes.

Havas, V., Taylor, J.S.H, Vaquero, L., de Diego-Balaguer, R., Rodríguez-Fornells, A., & Davis, M. H. (2018). Semantic and phonological schema influence spoken word learning and overnight consolidation. *Quarterly Journal of Experimental Psychology*, *71*(6), 1469–1481.

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), 406–417.

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), 572–592.

Hills, T. T., Maouene, J., Riordan, B., Smith, L. B., & , L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259–273.

Hino, Y., Pexman, P. M., & Lupker, S. J. (2006). Ambiguity and related- ness effects in semantic tasks: Are they due to semantic coding? *Journal of Memory and Language, 55*, 247–273.

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavioral Research Methods*, *45*(3), 718–730.

Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 385–402.

Hollis, G. (2020). Delineating linguistic contexts, and the validity of context diversity as a measure of a word's contextual variability. *Journal of Memory and Language*, *114*(June), 104146.

Horiba, Y., & Fukaya, K. (2015). Reading and learning from L2 text: Effects of reading goal, topic familiarity, and language proficiency. *Reading in a Foreign Language*, *27*(1), 22–46.

Horst, J. S. (2013). Context and repetition in word learning. *Frontiers in Psychology*, *4*(APR), 1–11.

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology*, *2*(FEB), 1–11.

Hsiao, Y., Bird, M., Norris, H., Pagán, A., & Nation, K. (2020). The influence of item-level contextual history on lexical and semantic judgments by children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(12), 2367–2383.

Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language*, *103*, 114–126.

Hulme, R. (2018). Incidental learning of new meanings for familiar words. [Unpublished doctoral dissertation]. University College London.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 113–125). London: Macmillan.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal*, *80*, 327–339.

Jacobs, G. M., Dufon, P., & Hong, F. C. (1994). L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading*, *17*(1), 19–28.

Jaeger, T. F. (2011, June 25). More on random slopes and what it means if your effect is not longer significant after the inclusion of random slopes [Blog post]. Retrieved from https://hlplab.wordpress.com/2011/06/25/more-on-random-slopes/

James, E., Gaskell, M. G., & Henderson, L. M. (2019). Offline consolidation supersedes prior knowledge benefits in children's (but not adults') word learning. *Developmental Science*, *22*(3), e12776.

Jastrzembski, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology*, *13*, 278–305.

Jenkins, J. R., Stein, M. L., &Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, *21*, 767-787.

Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin and Review*, *23*(4), 1214–1220.

Johns, B. T., Dye, M., & Jones, M. N. (2020). Estimating the prevalence and diversity of words in written language. *Quarterly Journal of Experimental Psychology*, *73*(6), 841-855.

Johns, B. T., Gruenenfelder, T. M., Pisoni, D. B., & Jones, M. N. (2012). Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *The Journal of the Acoustical Society of America, 132*(2), EL74–EL80.

Johns, B. T., Sheppard, C. L., Jones, M. N., & Taler, V. (2016). The role of semantic diversity in word recognition across aging and bilingualism. *Frontiers in Psychology*, *7*(MAY), 1–11.

Jones, M. N., Dye, M., & Johns, B. T. (2017). Context as an organizing principle of the lexicon. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 67, pp. 239–283). Amsterdam, the Netherlands: Elsevier.

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*(2), 115–124.

Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognitive Psychology*, *98*, 1–21.

Jorge-Botana, G., Olmos, R., & Sanjosé, V. (2017). Predicting Word Maturity from Frequency and Semantic Diversity: A Computational Study. *Discourse Processes*, *54*(8), 682–694.

Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role of context. *Journal of Experimental Child Psychology*, *166*(August 2017), 190–211.

Joseph, H. S., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition*, *133*(1), 238-248.

Jurafsky, D., & Martin, J. (2019). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd Edition Draft). Retrieved from https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

Kachergis, G., Shiffrin, R. M., Yu, C., & Shiffrin, R. M. (2009). Frequency and Contextual Diversity Effects in Cross-Situational Word Learning. *Proceedings of the Cognitive Science Society*, *31*(31), 2220–2225.

Kaefer, T., Neuman, S. B., & Pinkham, A. M. (2015). Pre-existing Background Knowledge Influences Socioeconomic Differences in Preschoolers' Word Learning and Comprehension. *Reading Psychology*, *36*(3), 203–231.

Kaivanpanah, S., & Rahimi, N. (2017). The Effect of Contextual Clues and Topic Familiarity on L2 Lexical Inferencing and Retention. *Porta Linguarum*, *27*, 47–61.

Kan, I. P., Alexander, M. P., & Verfaellie, M. (2009). Contribution of Prior Semantic Knowledge to New Episodic Learning in Amnesia. *Journal of Cognitive Neuroscience*, *21*(5), 938–944.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kost, C. R., Fost, P., & Lenzini, J. J. (1999). Textual and pictorial glosses: Effectiveness of incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals, 32*, 89-113.

Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Landi, N., Perfetti, C. A., Bolger, D. J., Dunlap, S., & Foorman, B. R. (2006). The role of discourse context in developing word form representations: A paradoxical relation between reading and learning. *Journal of Experimental Child Psychology*, *94*(2), 114–133.

Laufer, B. (1997). The lexical plight in second language reading; Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition*: A rationale for pedagogy (pp. 20–33). Cambridge, England: Cambridge University Press.

Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal*, *28*(1), 89–108.

Lev-Ari, S. (2016). How the Size of Our Social Network Influences Our Semantic Skills. *Cognitive Science*, *40*(8), 2050–2064.

Li, P., Burgess, C., & Lund, K. (2000). The acquisition of word meaning through global lexical co-occurrences. In E. V. Clark (Ed.), *Proceedings of the thirtieth Stanford child language research forum*. Stanford, CA: Center for the Study of Language and Information, 167–178.

Lin, C. & Huang, H. (2008). Meaning-inferred gloss and meaning-given gloss on incidental vocabulary learning. *Journal of National Taiwan Normal University: Humanities & Social Sciences, 53*(2), 87-116.

Lowell, R., & Morris, R. K. (2017). Impact of contextual constraint on vocabulary acquisition in reading. *Journal of Cognitive Psychology*, *29*(5), 551–569.

Mak, M. H. C. (2019). Why and how the co-occurring familiar object matters in Fast Mapping (FM)? Insights from computational models. *Cognitive Neuroscience*, *10*(4), 229–231. doi: 10.1080/17588928.2019.1593121

Mak, M. H., Hsiao, Y., & Nation, K. (2021). Lexical connectivity effects in immediate serial recall of words. *Journal of Experimental Psychology: Learning, Memory and Cognition 47(12), 1971–1997.* doi: 10.1037/xlm0001089

Mak, M. H. C. & Nation, K. (2019, September). Consolidation of novel words into semantic memory: the role of context. Poster presented at ESCoP 2019. http://escop2019.webs.ull.es/wp-content/uploads/2019/09/ESCoP2019-ABSTRACTS.pdf

Mak, M. H. C., & Twitchell, H. (2020). Evidence for preferential attachment: Words that are more well connected in semantic networks are better at acquiring new links in paired associate learning. *Psychonomic Bulletin & Review*, *27*, 1059-1069. doi: 10.3758/s13423-020-01773-0

Martin-Chang, S., Ouellette, G., & Bond, L. (2017). Differential Effects of Context and Feedback on Orthographic Learning : How Good Is Good Enough? Differential Effects of Context and Feedback on Orthographic. *Scientific Studies of Reading*, *21*(1), 17–30.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*, 375–407.

McDonald, S., & Ramscar, M. (2001). The Influence of Context on Judgements of Semantic Similarity. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. Merced.

McDonald, S., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, *44*(3), 295–323.

McKay, A., Davis, C., Savage, G., & Castles, A. (2008). Semantic Involvement in Reading Aloud: Evidence From a Nonword Training Study. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(6), 1495–1517.

McKeown, M.G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly, 20*(4), 482–496.

McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1155–1172.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *NIPS*, *13*, 3111–3119.

Millard, S. P. (2013). EnvStats: An R Package for Environmental Statistics. Springer, New York.

Miyasako, N. (2002). Does text-glossing have any effects on incidental vocabulary learning through reading for Japanese senior high school students? *Language Education & Technology*, *39*, 1–20.

Mondria, J. A., & Wit-De Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics*, *12*(3), 249–267.

Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological Science*, *26*(9), 1489–1496.

Murray, W.S., & Forster, K.I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, *111*, 721–756.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, *40*(5), 665–681.

Nagata, N. (1999). The effectiveness of computer-assisted interactive glosses. *Foreign Language Annals, 32*, 469–479.

Nagy, W.E. (1988). Teaching vocabulary to improve reading comprehension. Urbana, IL: NCTE, IRA, & ERICIRCS.

Nagy, W.E., & Anderson, R.C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, *19*, 304–330.

Nagy, W.E., & Herman, P. (1987). Depth and breadth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown, & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19–35). Hillsdale, NJ: Erlbaum.

Nagy, W. E., Herman, P. A., & Anderson, P. A. (1985). Learning words from context. *Reading Research Quarterly*, *20*, 233–253.

Nassaji, H. (2002). Schema theory and knowledge-based processes in second language reading comprehension: A need for alternative perspectives. *Language Learning, 52,* 439–481.

Nation, K. (2017). Nurturing a lexical legacy: reading experience is critical for the development of word reading skill. *Npj Science of Learning*, *2*(3), 1–4.

Nation, K., Angell, P., & Castles, A. (2007). Orthographic learning via self-teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of Experimental Child Psychology*, *96*(1), 71–84.

Nation, K., & Castles, A. (2017). Putting the learning in orthographic learning. In K. Cain, D. Compton, & R. Parrila (Eds.), *Theories of reading development* (pp. 147–168). Amsterdam, The Netherlands: John Benjamins.

Nelson, A. B., & Shiffrin, R. M. (2013). The Co-evolution of knowledge and event memory. *Psychological Review*, *120*(2), 356–394.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Ouellette, G. (2010). Orthographic learning in learning to spell: The roles of semantics and type of practice. *Journal of Experimental Child Psychology*, *107*(1), 50–58.

Ouellette, G., & Fraser, J. R. (2009). What exactly is a yait anyway: The role of semantics in orthographic learning. *Journal of Experimental Child Psychology*, *104*(2), 239–251.

Pagán, A., & Nation, K. (2019). Learning Words Via Reading: Contextual Diversity, Spacing, and Retrieval Effects in Adults. *Cognitive Science*, *43*(1), 1–24.

Pagán, A., Bird, M., Hsiao, Y., & Nation, K. (2020). Both Semantic Diversity and Frequency Influence Children's Sentence Reading. *Scientific Studies of Reading*, *24*(4), 356–364.

Paribakht, T., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 174–200). Cambridge, England: Cambridge University Press.

Paribakht, T., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition, 21,* 195–224.

Patterson, K., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia*, *30*, 1025–1040.

Paivio, A. (1971). *Imagery and Verbal Processes*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Paivio, A. (1991). Dual coding theory: retrospect and current status. *Canadian Journal of Psychology, 45*, 255–287.

Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. *EMNLP 2014*, 1532–1543.

Perea, M., Soares, A. P., & Comesaña, M. (2013). Contextual diversity is a main determinant of word identification times in young readers. *Journal of Experimental Child Psychology*, *116*(1), 37–44.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357–383.

Perfetti, C. A., & Roth, S. (1981). Some of the interactive processes in reading and their role in reading skill. In A. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 269–297). Hillsdale, NJ: Lawrence Erlbaum.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin and Review*, *15*(1), 161–167.

Pexman, P. M., Heard, A., Lloyd, E., & Yap, M. J. (2017). The Calgary semantic decision project: Concrete/abstract decision data for 10,000 English words. *Behavior Research Methods, 49*, 407–417.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology, 9*, 25–36.

Plummer, P., Perea, M., & Rayner, K. (2014). The influence of contextual diversity on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *40*(1), 275–283.

Plunkett, K. (1998). Language Acquisition and Connectionism. *Language and Cognitive Processes*, *13*(2-3), 97–104.

Pulido, D. (2003). Modeling the role of second language proficiency and topic familiarity in second language incidental vocabulary acquisition through reading. *Language Learning*, *53*, 233–284.

Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, *54*(3), 469–523.

Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. *Applied Linguistics*, *28*(1), 66–86.

Qiu, M., & Johns, B. T. (2020). Semantic diversity in paired-associate learning: Further evidence for the information accumulation perspective of cognitive aging. *Psychonomic Bulletin & Review*, *27*(1), 114–121.

R Core Team. (2019). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ramscar, M., Sun, C. C., Hendrix, P., & Baayen, H. (2017). The Mismeasurement of Mind: Life-Span Changes in Paired-Associate-Learning Scores Reflect the "Cost" of Learning, Not Cognitive Decline. *Psychological Science*, *28*(8), 1171–1179.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.

Recht, D. R., & Leslie, L. (1988). Effect of Prior Knowledge on Good and Poor Readers' Memory of Text. *Journal of Educational Psychology*, *80*(1), 16–20.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*, 274–280.

Reichle, E. D., & Perfetti, C. A. (2003). Morphology in word identification: A word experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading*, *7*, 219–237.

Rodd, J. (2019, February 27). How to Maintain Data Quality When You Can't See Your Participants. *Observer*. https://www.psychologicalscience.org/observer/how-to-maintain-data-quality-when-you-cant-see-your-participants

Rodd, J. (2020). Settling Into Semantic Space: An Ambiguity-Focused Account of Word-Meaning Access. *Perspectives on Psychological Science*, *15*(2), 411-427.

Rodd, J., Gaskell, M. G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*(2), 245–266.

Rodd, J., Gaskell, M. G., & Marslen-Wilson, W. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science*, *28*(1), 89–104.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255.

Roodenrys, S., Hulme, C , Alban, J., Ellis, A.W., & Brown, G.D.A. (1994). Effects of word frequency and age of acquisition on short-term memory span. *Memory and Cognition, 22*, 695–701.

Rosa, E., Tapia, J. L., & Perea, M. (2017). Contextual diversity facilitates learning new words in the classroom. *PLoS ONE*, *12*(6), 1–12.

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), 12663–12668.

Rumelhart, D.E., & Ortony, A. (1977). The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 99–135). Hillsdale, NJ: Erlbaum.

Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System, 6*, 70–78.

Saussure, F. de. (1916). *Cours de linguistique générale*. Paris: Payot. (English translation: *A course in general linguistics*. New York: Philosophical Library).

Scha, R.J.H., Bruce, B.C., & Polanyi, L. (1986). Discourse understanding (Tech. Rep. No. 311). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Schwanenflugel, P., & LaCount, K. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(2), 344–354.

Scott, R. M., & Fisher, C. (2012). 2.5-Year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*(2), 163–180.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*, 151–218.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72*, 95–129.

Shefelbine, J. L. (1990). Student factors related to variability in learning word meanings from context. *Journal of Literacy Research*, *22*(1), 71–97.

Siew, C. S. Q. (2019). spreadr: An R package to simulate spreading activation in a network. *Behavioral Research Methods*, *51*(2), 910–929.

Siew, C. S. Q., & Vitevitch, M. S. (2016). Spoken word recognition and serial recall of words from components in the phonological network. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(3), 394–410.

Siew, C. S. Q., & Vitevitch, M. S. (2020). An investigation of network growth principles in the phonological language network. *Journal of Experimental Psychology: General. 149*(12), 2376–2394.

Silge, J. (2017). *Word Vectors with tidy data principles* [Blogpost]. Retrieved from https://juliasilge.com/blog/tidy-word-vectors/

Silge, J. & Robinson, D. (2016). tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *JOSS*, *1*(3).

Simonsen, S., & Singer, H. (1992). Improving reading instruction in the content areas. In S.J. Samuels & A.E. Farstrup (Eds.), *What research has to say about reading instruction* (pp. 200–219). International Reading Association.

Spilich, G.J., Vesonder, G.T., Chiesi, H.L., & Voss, J.F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavioral*, *18*, 275–290.

Stanovich, K. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*, 32–71.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*(1), 41–78.

Suárez-Coalla, P., & Cuetos, F. (2017). Semantic and phonological influences on visual word learning in a transparent language. *Quarterly Journal of Experimental Psychology*, *70*(4), 772–781.

Sulzby, E. (1985). Children's emergent reading of favorite storybooks: a developmental study. *Reading Research Quarterly*, *20*(4), 458–481.

Tamminen, J., Davis, M. H., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology*, *79*, 1–39.

Tamura, N. (2015). What's in a Story? Children's Learning of New Written Words via Reading Experience. [Unpublished doctoral dissertation]. University of Oxford.

Taylor, J. S. H., Duff, F. J., Woollams, A. M., Monaghan, P., & Ricketts, J. (2015). How Word Meaning Influences Word Reading. *Current Directions in Psychological Science*, *24*(4), 322–328.

Taylor, J. S. H., Plunkett, K., & Nation, K. (2011). The Influence of Consistency, Frequency, and Semantics on Learning to Read: An Artificial Orthography Paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(1), 60–76.

Teng, F. (2019). The effects of context and word exposure frequency on incidental vocabulary acquisition and retention through reading. *Language Learning Journal*, *47*(2), 145–158.

van den Broek, G. S. E., Takashima, A., Segers, E., & Verhoeven, L. (2018). Contextual Richness and Word Learning: Context Enhances Comprehension but Retrieval Enhances Retention. *Language Learning*, *68*(2), 546–585.

van Dongen, E. V., Thielen, J. W., Takashima, A., Barth, M., and Fernández, G. (2012). Sleep supports selective retention of associative memories based on relevance for future utilization. *PLoS One, 7*, e43426.

Wamsley, E. J., Hamilton, K., Graveline, Y., Manceor, S., & Parr, E. (2016). Test expectation enhances memory consolidation across both sleep and wake. *PLoS ONE*, *11*(10), 1–15.

Wang, H. C., Castles, A., Nickels, L., & Nation, K. (2011). Context effects on orthographic learning of regular and irregular words. *Journal of Experimental Child Psychology*, *109*(1), 39–57.

Wang, H., Savage, G., Gaskell, M. G., Paulin, T., Robidoux, S., & Castles, A. (2017). Bedding down new words: Sleep promotes the emergence of lexical competition in visual word recognition. *Psychonomic Bulletin & Review*, *24*(4), 1186–1193.

Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, *1*, 287–307.

Webb, S. (2007). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research*, *11*(1), 63–81.

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*(2), 232–245.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, *1*, 59–85.

Wilhelm, I., Diekelmann, S., Molzow, I., Ayoub, A., Molle, M., & Born, J. (2011). Sleep Selectively Enhances Memory Expected to Be of Future Relevance. *Journal of Neuroscience*, *31*(5), 1563–1569.

Wilkinson, K. S., & Houston-Price, C. (2013). Once upon a time, there was a pulchritudinous princess . . .: The role of word definitions and multiple story contexts in children's learning of difficult vocabulary. *Applied Psycholinguistics*, *34*(3), 591–613.

Williams, R., & Morris, R. K. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, 16(1/2), 312–339.

Williams, S. E., & Horst, J. S. (2014). Goodnight book: Sleep consolidation improves word learning via storybooks. *Frontiers in Psychology*, *5*, 184.

Wittgenstein, L. (1953). *Philosophical Investigations*. (Translated by Anscombe, G.E.M.). Blackwell.

Wong, M. Y., & Underwood, G. (1996). Do bilingual children read words better in lists or in context? *Journal of Research in Reading*, *19*(1), 61–76.

Wu, J. (2019). chinese.misc: Miscellaneous Tools for Chinese Text Mining and More. R package version 0.2.1. https://CRAN.R-project.org/package=chinese.misc

Yanek, D. (2018, December 12). Why do toddlers want to read the same book over and over again? [Blog Post]. Retrieved from https://www.todaysparent.com/toddler/toddler-behaviour/why-toddlers-want-to-read-the-same-book/

Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review, 18*, 742–750.

Yeh, Y., & Wang, C.-W. (2003). Effects of multimedia vocabulary annotations and learning styles on vocabulary learning. *CALICO Journal, 21*, 131–144.

Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, *10*, 85–101.

Appendices

Appendix 1A

List of base words in Experiments 1 to 4

Experiment 1	Experiment 2	Experiments 3 and 4
avidity	*avidity*	*avidity*
castigate	*castigate*	*castigate*
constellation	*constellation*	*constellation*
diffident	*diffident*	*diffident*
latent	*latent*	*latent*
mendacious	*mendacious*	*mendacious*
predilection	*predilection*	*predilection*
proliferation	*proliferation*	*proliferation*
shenanigans	*shenanigans*	*shenanigans*
vagarious	*vagarious*	*vagarious*
	architecture	*architecture*
	imbroglio	*imbroglio*
		abstruse
		comity
		extirpate
		fastidious

Appendix 1B

Pseudowords (used to replace the base words) across Experiments 1 to 4

Experiments 1 and 2	
List 1	**List 2**
allitive	abbusion
appurate	attusive
arregory	chesume
dratula	chrabble
flematic	draimant
genisty	glansact
lainstrom	lonbust
meciancy	mauranty
ossanic	prasmodic
plurious	splange
spredule	bropence
chyrene	chratum

Note. Pseudowords in the shaded boxes were used in Experiment 2 only.

Experiments 3 and 4
alimethy
allitive
appurate
arregory
champede
chyrene
dratula
flematic
genisty
lainstrom
meciancy
notaruse
oddent
ossanic
plurious
spredule

Appendix 1C

Semantic associates of the base words

Base words	Associate 1	Associate 2
1. *Avidity*	Enthusiasm	Passion
2. *Castigate*	Criticise	Reprimand
3. *Constellation*	Variety	Group
4. *Diffident*	Shy	Timid
5. *Latent*	Conceal	Hidden
6. *Mendacious*	Dishonest	Fraudulent
7. *Shenanigans*	Mess	Problem
8. *Predilection*	Inclination	Preference
9. *Proliferation*	Increase	Expansion
10. *Vagarious*	Erratic	Unpredictable
11. *Architecture*	Construction	Structure
12. *Imbroglio*	Humiliation	Embarrassment
13. *Abstruse*	Obscure	Complicated
14. *Comity*	Unity	Solidarity
15. *Extirpate*	Eliminate	Destroy
16. *Fastidious*	Careful	Precise

Note.

Items 1-10 were used in Experiment 1.

Items 1-12 were used in Experiment 2.

Items 1-16 were used in Experiment 3 and 4.

Appendix 1D

Parametric settings for the spreadr simulations in Simulations 1 and 2

1. Start run: 50 (Simulation 1); 40 (Simulation 2a); 60 (Simulation 2b)

Start run refers to how much activation a particular node, usually the word of interest, initially receives at Time 0. In Siew's (2019) study, an initial activation level of 100 was used to simulate lexical retrieval, false memory, and semantic priming. The lexical items in these paradigms were highly familiar words (e.g., *sleep*), so their mental representations must be fairly accurate and robust. When activated, these known words should have significantly higher activation levels than the newly learned words in Experiments 1 and 2. Therefore, the initial activation level for the novel words was reduced to 50 in Simulation 1 (where each target word had a frequency of 6), to 40 in Simulation 2a (where each target word had a frequency of 5), and to 60 in Simulation 2b (where each target word had a frequency of 8).

2. Retention: 0.1 to 0.9

Retention refers to the percentage of activation that remains in a node at each time step. Following Vitevitch et al. (2011) and Siew (2019), nine values were used: 0.1 to 0.9, in increments of 0.1. This was to ensure that "the simulation results were consistent across various values of retained level" (Siew, 2019).

3. Decay: 0.1

Decay refers to the proportion of "lost" activation at each time step of simulation. This parameter was not manipulated in Siew's (2019) simulation and was set at 0. However, given there exists empirical work (e.g., McKoon & Ratcliff, 1992) suggesting that spreading activation is a limited resource that decays over time, the current simulations attempted take this decaying property into account by setting the decay level to an arbitrary value of 0.1.

4. Suppress: 0.05

Suppress refers to the threshold where activation is suppressed to 0. This parameter instantiates the notion that nodes with extremely low activation levels are non-active and non-influential in spreading activation process (Siew, 2019). In her simulations, Siew kept this at 0, but this may not be appropriate for Simulations 1 and 2. Because firstly, there are significantly more nodes in the current networks (>200) than Siew's (~20), and secondly,

some nodes in the current networks, especially those that only co-occurred with the target word once in the reading passages, are expected to have very little or no influence on the spreading activation process. Therefore, the “suppress” parameter was arbitrarily set at 0.05.

5. Time step: 10

This parameter refers to the number of time steps over which the spreading activation process occurs. Following Vitevitch et al. (2011) and Siew (2019), this was set at 10.

Appendix 5A

The 32 neutral sentences used in Experiments 3 and 4

Base word	Neutral sentences
Abstruse	The connection was abstruse but strong.
	He professed no knowledge, so he avoided abstruse inquiry.
Architecture	If you want to fully understand the software architecture, read this.
	The genetic architecture of anorexia is poorly understood.
Avidity	Given Jake's avidity for sunshine, this is a good location.
	The upper-class has a avidity for eccentricity.
Castigate	It is easy to castigate people you don't like.
	You cannot just castigate the journalists for doing their job.
Comity	A society needs comity to prosper.
	He wishes comity for the country.
Constellation	The metabolic syndrome is a constellation of problems that include obesity.
	A constellation of circumstances seemed to help drive her to the final.
Diffident	She was diffident about stating her opinion.
	I thought he was diffident but not everyone agreed.
Extirpate	We will never be able to extirpate anti-Christian bigotry.
	During the Cold War, mutually hostile ideologies sought to extirpate each other.
Fastidious	You need to be fastidious if you wish to succeed.
	He was a fastidious fireman.
Imbroglio	Jennie found herself in a imbroglio and did not know what to do.
	Dealing with a imbroglio is not easy.
Latent	The problems have remained latent for several years.
	There is some latent, if wary, gratitude for American aid.
Mendacious	I wish I knew he has an mendacious personality!
	I find mendacious people despicable.
Predilection	Peter is a young lad with a predilection for telling tales.
	She wished she didn't have a deep predilection for wine.
Proliferation	The cell proliferation ability of MG-63 was astounding.
	There is a proliferation of fake drug in the market.
Shenanigans	The win did not happen without shenanigans, so shame on them.
	She has never been accused of any shenanigans in her career.
Vagarious	The scientist is surprised at the vagarious behaviours of the particle.
	The competition can become vagarious, so keep that in mind.

Appendix 6A

The definitions used in the with-definition group of Experiment 4

Base words	Definitions	List
Abstruse	difficult to understand	1
Architecture	complex structure	1
Avidity	keen enthusiasm	2
Castigate	criticize someone severely	1
Comity	harmonious behaviour	1
Constellation	a group of something	2
Diffident	modest or shy	1
Extirpate	eradicate completely	2
Fastidious	concerned about details	1
Imbroglio	a confusing / distressing situation	1
Latent	hidden or concealed	2
Mendacious	not telling the truth	2
Proliferation	rapid increase	2
Predilection	special liking for something	2
Shenanigans	secret activity	1
Vagarious	very unpredictable	2