Entrenchment Revisited: Some Old and New Concepts and Their Empirical Validation

Entrenchment Revisited: Alte und Neue Konzepte und Ihre Empirische Validierung

der Philosophischen Fakultät und Fachbereich
Theologie der Friedrich-Alexander-Universität
Erlangen-Nürnberg
zur
Erlangung des Doktorgrades
Doktorin der Philosophie (Dr.Phil.)

vorgelegt von

Evelin Balog

Als Dissertation genehmigt
von der Philosophischen Fakultät und Fachbereich Theologie
der Friedrich-Alexander-Universität Erlangen-Nürnberg
Tag der mündlichen Prüfung: 30.03.2023

Vorsitzender der Promotionsorgan: Prof. Dr. Thomas Demmelhuber

Gutachter/in: Prof. Dr. Thomas Herbst
Prof. Dr. Ewa Dąbrowska

Zusammenfassung in deutscher Sprache

Das Konzept des Entrenchment steht seit über drei Jahrzehnten im Mittelpunkt der kognitiven Linguistik. Es hat sich von einem rein theoretischen zu einem empirisch untersuchten Konzept entwickelt. Betrachtet man die vorhandenen Definitionen von Entrenchment, so wird deutlich, dass die meisten von ihnen die beiden grundlegenden Komponenten enthalten, die ursprünglich von Langacker (1987: 59) vorgeschlagen wurden: Wiederholung ("Every use of a [linguistic] structure has a positive impact on its degree of entrenchment") und ganzheitliche (holistic) Verarbeitung/Speicherung der verfestigten sprachlichen Einheit ("…a novel structure becomes progressively entrenched, to the point of becoming a unit"). Diese beiden Komponenten deuten darauf hin, dass der häufige Kontakt mit der Sprache einen Prozess auslöst, der zur Einprägung und Speicherung der sprachlichen Einheiten führt.

Die vorliegende Dissertation befasst sich mit der Beziehung zwischen sprachlichen, sozialen und kognitiven Faktoren und Entrenchment von sprachlichen Elementen in den Köpfen (minds) der Sprecher:innen. Dieses weit gefasste Thema lässt sich in die folgenden Hauptfragen untergliedern:

- Was genau ist mit Entrenchment gemeint und wie kann es operationalisiert und empirisch getestet werden?
- Welche Faktoren tragen zum Entrenchment bei, wie modulieren diese Faktoren den Grad des Entrenchments und was sagen Unterschiede im Grad des Entrenchments über den Prozess selbst aus?
- Inwieweit ist der Grad des Entrenchments Teil des Wissens eines Sprechers und können wir dieses Wissen nutzen, um ein tieferes Verständnis des Konzepts Entrenchment zu erlangen?

Mithilfe eines auf einer Kombination verschiedener Methoden beruhenden Forschungsdesigns werden diese drei Hauptfragen kritisch bewertet und empirisch untersucht. Durch die Zusammenführung von Erkenntnissen aus der Linguistik, der Psycholinguistik, der Neurologie und der kognitiven Psychologie zum Thema Entrenchment zielt die vorliegende Dissertation darauf ab, ein realistisches Verständnis von Entrenchment, Gedächtnis und Automatizität zu vermitteln und zu zeigen, wie sie ein dynamisches System bilden, das für den lebenslangen kognitiven Reorganisationsprozess im mentalen Lexikon/Konstruktikon von Sprecher:innen.

Die vier Experimente und die Fallstudie (case study), die in dieser Dissertation vorgestellt werden, untersuchen die Beziehung zwischen Frequenz, Vertrautheit, Lesegewohnheiten, korpusgestützten Assoziationsmasse und Entrenchment. Die Studien befassten sich mit verschiedenen sprachlichen Ebenen; die ersten beiden Experimente konzentrierten sich auf die Wortebene, die anderen beiden auf in Sätze eingebettete Adjektiv-Nomen-Wortkombinationen. In den Experimenten wurden verschiedene psycholinguistische Methoden eingesetzt, darunter Rapid Word Naming, Sentence Reading and Verbatim Recall, um Entrenchment empirisch zu validieren. Es wurden über 120 Teilnehmer getestet.

Das Hauptexperiment konzentriert sich auf Automatizität und die Stärke der Gedächtnisrepräsentationen. Ziel war es, in allen Experimenten einen Zusammenhang zwischen den Faktoren, den Prozessen und dem Produkt des Entrenchments herzustellen. Automatizität wird im Rahmen dieser Dissertation als das Produkt des Entrenchments operationalisiert. Automatizität kann viele Formen annehmen. In allen vier Experimenten stand die zeitliche und phonetische Reduktion im Mittelpunkt der Untersuchung. Diese beiden Reduktionsprozesse werden als Manifestationen der Automatizität in der Sprachproduktion behandelt. Die Experimente haben miteinbezogen die kognitiven Effekte und die psychologischen Effekte. Da die ersten beiden Experimente weniger umfangreich sind, können sie nicht als Eckpfeiler der Validierung oder Falsifizierung von Entrenchment betrachtet werden, sondern eher als ergänzende Experimente, die unser Verständnis des Phänomens des Entrenchments vertiefen und die Entwicklung von Tests unterstützen sollen, die die gesamte Komplexität des Phänomens erfassen können.

Die Ergebnisse der Experimente zeigen, dass die Wirkung der linguistischen Faktoren, wie Häufigkeit und Übergangswahrscheinlichkeit (transitional probability), stark von den Prozessen abhängt, die sie vorhersagen sollen, d. h. Verstehen, Produktion und Lernen. Verschiedene Sprachbenutzer können auch eine unterschiedliche Sensibilität für diese Prädiktoren zeigen und sich je nach Aufgabe und Kommunikationssituation mehr oder weniger stark auf sie verlassen. Darüber hinaus haben die Ergebnisse gezeigt, dass sich die Wirkung verschiedener linguistischer Faktoren addieren kann, sodass sich sogar noch höhere förderliche Effekte ergeben. Schließlich lieferte die abschließende Studie die empirische Bestätigung für den zu Beginn der Dissertation skizzierten theoretischen Rahmen, d. h., dass Automatizität als Endprodukt des Entrenchments angesehen werden kann, da die Stärke das Gedächtnis–repräsentationen einen großen Einfluss darauf hat, ob der Prozess mehr oder weniger automatisch abläuft. Zusammenfassend lässt sich sagen, dass die hier vorgelegten Studien die Hypothesen stützen, die sich aus dem gebrauchsbasierten Ansatz ergeben, und dass die Erkenntnisse, die aus der Analyse der Interaktion zwischen den sozialen, kognitiven und linguistischen Prädiktoren gewonnen wurden, dazu ermutigen, den gebrauchsbasierten Rahmen zu verfeinern, indem untersucht wird, wie diese Faktoren zu den verschiedenen Prozessen beitragen und wie die kognitiven Fähigkeiten der Sprachbenutzer den Grad ihrer mentalen Repräsentationen prägen.

Content

List of tables

List of figures

Conventions

List of Predictor Variables and their Abbreviations

ASR	average speech rate
PronLength	pronunciation length of the target words in seconds
LogFreqNoun	log-transformed frequency of the noun, if not indicated otherwise, it was extracted from the GermanWeb
WordLength	refers to the length of the word in letters
TP	transitional probability
TP forward(sc)	forward transitional probability. _sc means that the probability was scaled
TP_backward(sc)	backward transitional probability
TP_compound	transitional probability of the compound. The likelihood that one of the free morphemes is followed by the other free morpheme
LogRelFreqBase	log-transformed relative frequency of the base
FreqDWDS	frequency extracted from the DWDS, if not specified otherwise, it refers to the KernKorpus (1900-1999)
FreqWeb	frequency extracted from the GermanWeb
FreqBase	frequency of the base
FreqStemEnding	frequency of the ending of the compound
LogFreqWeb	log-transformed frequency extracted from the GermanWeb
MI	mutual information
CD	contextual diversity
NCD	noun contextual diversity
MentProcRT	mental processing speed
MentProcAcc	mental processing accuracy
FreqModifier	frequency of the modifier (adjective)
SentRatePhon	average speech rate of the sentence measured in phon/sec
FreqColloc	frequency of the collocation
FreqSchwaWord	frequency of the free morpheme containing the pre-nasal schwa
FreqSecMorph	frequency of the free morpheme in the polymorphic compound that does not have a pre-nasal schwa
ERP	event-related potentials. It is measured using (EEG) electroencephalography

1 Introduction

Language is one of the most complex of the human cognitive abilities. Even before birth, we start learning language and continue this learning process throughout our lives. At birth, babies can already distinguish between familiar and unfamiliar sounds. In the process of becoming more and more proficient in a language, we rely more on our cumulative knowledge of the language to predict what someone is going to say or which words are generally likely to occur in particular contexts. Suppose several people are exposed to the sentence fragment: *I intended to give you the card on Saturday, but I completely*... It is highly probable that most of them, provided that they are proficient in English, would end the sentence with the word *forgot*. The question is, how do they know that the missing word is "*forgot*"? Is it because this adverb-verb combination frequently occurs in the language? Or is it because speakers are just familiar with the collocation *completely forget*? Could it be because the adjective *completely* is often followed by the verb *forget*? Another explanation might be that this combination is stored as one holistic unit, a.k.a. chunk, or perhaps even that the context itself limits the possible endings. Research has found evidence to support each of these possible explanations. Generally, we can claim that they know which word is missing because it is entrenched in their mental lexicon. Entrenchment is a widely discussed phenomenon in cognitive and usage-based linguistics. There is a strong connection between entrenchment and the usage-based approach since both of these approaches assign an essential role to language usage. The usage-based approach integrates psycholinguistic and sociolinguistic perspectives by treating linguistic knowledge as dynamic networks of constructions that are shaped by social context, thus accounting for productivity, real-time processing, and learnability. Within this framework, mental representations consist of form-meaning pairings that emerge and are constantly shaped and updated by language usage until they reach the stage of automatic processing and production (Goldberg 1995, 2006; Tomasello 2005; Bybee 2006). Ultimately, language use continuously shapes mental representations of language, leading to linguistic constructions being entrenched to varying degrees depending on the speaker's usage-based experience of each particular construction. This framework is theoretically underpinned by the fact that speakers prefer to use expressions they have already heard/seen and used themself (Bybee 2010). Speakers process these linguistic items faster (Arnon & Clark 2011; Arnon & Snider 2010) and produce them more fluently (Verhagen 2020; Yoshimura & MacWhinney 2007), which implies that these expressions have a unique mental representation, and as such, they are easier to access than unusual and entirely new sequences. In other words, entrenchment can be generally defined as the development of routine behaviours. These routine behaviours develop quite generally and in any context (i.e. linguistic and non-linguistic) when, through frequent repetition, motor activities become 'chunked' or welded into one longer sequence (Langacker 2000: 3:4). These motor activities can be observed in activities such as learning to type on a keyboard or riding a bike but are also evident in the motor activities needed to pronounce words. Additionally, this is a further possible explanation for the fact that most language learners know that *renewable* is followed by *energy*, *helpful* by *hand*. As a result of chunking, language users do not have to assemble these multi-word combinations but can access them together as a holistic unit.

The present dissertation deals with the relationship between the linguistic, social and cognitive factors and the entrenchment of linguistic items in the minds of language users. This broad issue can be broken down into the following main questions: Firstly, what exactly is meant by entrenchment, and how can it be operationalised and empirically tested? Second, which predictors contribute to entrenchment? Additionally, how do these predictors modulate the degree of entrenchment, and what can variation in the degrees of entrenchment tell us about the process itself? Third, to what extent is the degree of entrenchment part of a speaker's knowledge? Furthermore, can we use this knowledge to gain a deeper understanding of the concept of entrenchment? Using a multi-method research design, these three main questions will be critically evaluated and empirically investigated. By bringing together the findings on entrenchment from the field of linguistics, psycholinguistics, neurology, and cognitive psychology, the present dissertation aims to provide a realistic understanding of entrenchment, memory, and automaticity and how they form a dynamic system that is responsible for the lifelong cognitive reorganisation process that takes place in the speakers' mind. The present dissertation is divided into four main parts.

The first part will discuss how linguistic predictors trigger a range of learning processes that subsequently lead to the entrenchment of linguistic items and structures. Moreover, it will be aimed to develop a dynamic model of entrenchment that incorporates the linguistic factors that stem from the linguistic input and the cognitive skills that allow language users to register the input, encode it into memory and strengthen the representation of the linguistic items. The aim of this dynamic model will be to divide a concept of entrenchment into smaller empirically measurable components. The role of each component in the process of entrenchment will be discussed in detail. This section is further subdivided into three parts. The first subsection discusses the linguistic predictors of entrenchment and has the objective to highlight how each of them facilitates entrenchment and outlines the idea that these predictors should be analysed as a joint effect rather than separating them into single standalone effects. The second subsection discusses the processes that are related to entrenchment. In the initial part of the section, three different learning processes will be discussed, and the quantitative changes these learning processes trigger will be highlighted. This subsection primarily focuses on the cognitive processes that enable speakers to entrench the different linguistic structures, thereby emphasising the necessity to analyse individual differences that determine how quickly and easily different speakers entrench the different structures. The final section in this part deals with the end product of entrenchment - automaticity. First, the concept of automaticity will be outlined, after which its different manifestations in language comprehension and production will be discussed.

The second part of the dissertation will focus on the experimental part of this study. An important part of this section is to emphasise that the phenomena under consideration require a multi-method approach to be investigated and to provide the readers with the relevant background knowledge on the different psycholinguistic experimental paradigms and outline why each specific task was selected to test entrenchment. The different tests that are the backbone of the four experiments will be discussed in detail. It will be followed by the description of the participants and the procedure of each experiment. Finally, the data analysis is described in detail, providing the result of each statistical model.

The third part of the dissertation aims to summarise the main findings of the four experiments and is divided into three parts. Each part represents one of the main findings and is followed by a discussion of the results.

The closing chapter will first offer a summary of this work before discussing the theoretical implications of the finding. It is hoped that the discussion provided in this section will contribute to the ongoing debates in cognitive linguistics about the role that different linguistic predictors play in entrenchment, the connection between memory strength and automaticity, and the benefit of relying on the metalinguistic judgements obtained from language users. Finally, some of the limitations of the study and some future research directions will be outlined.

2 Operationalising entrenchment

The concept of entrenchment has been the focus of cognitive linguistics for over three decades. It evolved from a purely theoretical into an empirically investigated concept. Looking at the existing definitions of entrenchment, it becomes clear that most of them incorporate the two fundamental components originally proposed by Langacker (1987: 59): repetition ("Every use of a [linguistic] structure has a positive impact on its degree of entrenchment") and holistic processing/storage of the entrenched linguistic unit ("…a novel structure becomes progressively entrenched, to the point of becoming a unit"). These two components indicate that frequent exposure to language triggers a process that leads to the memorisation and storage of linguistic units.

The main trigger of this process is repetition. Generally, repetition is defined as the frequency of occurrence of a linguistic unit. The discovery of the frequency effect by (Cattell 1886) is one of the most significant linguistic breakthroughs and, since then, has primarily shaped how we conceptualise language processing and production. The concept of entrenchment is predominantly associated with processing and production advantages of high-frequency words, collocations and constructions, and that is why it is of no surprise that eventually, the effects of frequency and entrenchment merge together, as both of them are associated with similar manifestations - namely, fast, effortless (Bybee 2007a; Ambridge et al. 2015) and more accurate processing (Balota et al. 2012) and resilience to brain damage and ageing (Rogers & McClelland 2004). In other words, with time, frequency has not only become defined as the driving force behind entrenchment but as entrenchment itself. This premise has its foundation in the corpus-to-cognition theory. The assumption that entrenchment can be defined entirely through corpus-derived frequency resides in the belief that corpora can provide an accurate picture of what language users are actually exposed to. However, no one would deny that people are exposed to different kinds of language depending on their social environment. Hence, it is realistic to assume that "no two speakers have the same language, because no two speakers have the same experience of language" (Hudson 1996: 11).

The second fundamental component is the process that leads to changes in mental representations. Most definitions focus on the processes that are associated with or lead to entrenchment, such as the strengthening of knowledge (Hilpert & Diessel 2017; Theakston 2017; Blumenthal-Dramé 2012, 2017), automatisation

(Langacker 1987; Bybee 2002b, 2013; Schmid 2017a; Schmid 2020; Hartsuiker & Moors 2017), faster and more effortless processing (Bybee 2007a; Divjak & Caldwell-Harris 2015; Divjak 2019), memory consolidation (Blumenthal-Dramé 2012; Bakker et al. 2015; Divjak 2019), and chunking (De Smet 2017; Blumenthal-Dramé 2012; Langacker 1987). The processes specified above are triggered by a range of linguistic factors, including frequency, and fall into two sets that emphasise both the quantitative and qualitative changes resulting from entrenchment. The first set focuses on mental representation strengthening and memory consolidation and chunking, which impact the organisation of an individual's linguistic system. These processes determine how linguistic memories are formed, stored, and retrieved. Memories can be defined as the reactivation of specific groups of neurons formed from persistent changes in the strength of connections between neurons. The single neurons can represent the sound or syllables of the language, and the particular sequence in which they are activated determines which linguistic unit is retrieved. The more often the neurons are activated in that particular sequence, the stronger the memory will be (Dehaene et al. 2015). Suppose these linguistic units are made up of smaller sub-entities (e.g., compounds, different phrasal verbs, collocations, and fixed expressions). In that case, with time, they start to be perceived as one unit because the neurons responsible for the activation of each component merge into one neural circuit that represents all the components of the structure. Thus, these units will be retrieved from memory in one single step from that point on. This interpretation implies a threshold level, and upon reaching it, sequences of linguistic units that frequently co-occur are reanalysed as one single precompiled unit (Blumenthal-Dramé 2012: 67).

The second set of features, namely automatisation and faster and more effortless processing, are related to changes in the use of stored entities rather than their inventory. They represent the qualitative changes. These qualitative changes signify the different degrees of entrenchment. The degree of entrenchment depends not only on linguistic factors such as frequency but also on the cognitive skills of language speakers that enable them to process frequency and create new memories. Cattell (1886: 65), in his seminal paper, described automaticity in the following way:

> "… we can recognise a single colour or picture in a slightly shorter time than a word or letter, but take longer to name it. This is because in the case of words and letters the association between the idea and name has taken place so often that the process has become automatic, whereas, in the case of colours and pictures, we must by a voluntary effort choose the name."

It is apparent from the quote above that Cattell (1886), even prior to Langacker (1987), highlighted the role of repetition and how repeated actions become automatic with time. The associations Cattell (1886) refers to are nothing other than the connections between neurons discussed above. The strengthening of these associations leads to decreasing the activation threshold, resulting in faster and more effortless processing and naming. These qualitative changes are the final stage of entrenchment.

Within the framework of this dissertation, entrenchment is defined as a threefold concept triggered by usage and exposure frequency that leads to the strengthening and formation of mental representations and their reorganisation into chunks, resulting in effortless and fluent processing and production of the entrenched units. The three components by which entrenchment is defined include the linguistic factors that are the driving forces of entrenchment (frequency, familiarity, contextual diversity, and transitional probability), the processes that lead to the encoding of these linguistic items into memory (statistical learning, Hebbian learning, memory consolidation, chunking), and the end product of this process: automaticity. In other words, entrenchment is

I) a cognitive process and state at the same time that is highly influenced by language users' communicative habits and should significantly affect language acquisition, processing, and production:

II) a manifold concept that involves the processing the linguistic units, their encoding and storage in memory;

III) a gradual process influenced by several non-language-related predictors like the speaker's social environment and cognitive abilities.

The three-folded categorisation of entrenchment allows for breaking down this complex phenomenon into more measurable subunits. Entrenchment cannot take place without frequently occurring language input. However, focusing only on the linguistic factors that are the driving forces behind the process would suggest that entrenchment is a property of words because their frequency of occurrence is considered, by some linguists, to be a direct measure of the degree of entrenchment (Stefanowitsch & Flach 2017: 105). Entrenchment cannot be a property of the word because this would suggest that frequent words are equally entrenched in the minds of every speaker of the language. This claim would completely neglect the existence of any individual and social differences. In other words, frequently observed linguistic items can only become entrenched if and only if the language users register and store them in their memory. Corpus data can only vaguely represent which items are stored in the mind of the speakers through the lens of their language usage; however, it shows only a tendency and cannot provide a precise representation of which items and to what degree units are entrenched in the mind of a particular language user. Moreover, frequency obtained from any corpora is a relatively stable measure of occurrence (it can increase if the particular corpus is constantly updated with newer occurrences; however, it is very unlikely that it will decrease), unlike the mental representations of the language users, which are continuously updated as a result of the speakers' everyday language use. The strengthening of the representation of an item is just as likely as that it fades away. Consequently, the social and cognitive factors that allow speakers to demonstrate the facilitatory effects of these measures are just as vital as the driving forces themselves.

2.1 The components of entrenchment

The vast majority of the studies dealing with entrenchment focused on the driving forces, defined in this dissertation as the linguistic factors that measure prior exposure to language and the history of language use. The primary focus has been devoted to frequency, which is not surprising as the original definition by Langacker (1987:59) emphasises the inevitable role frequency plays in entrenchment. With more advanced experimental and statistical methods, the scope of linguistic factors has been widened with measures that take into account not only the pure frequency of individual words but also the context in which these words occur.

Shifting the attention from frequency to context-based measures opened up a whole new field of research. The tendency that emerges from the existing literature is that the linguistic factors that trigger entrenchment are in competition. Complex statistical models are used to determine which of the driving forces is the most significant or can explain the highest portion of the variation in the data. In the following sections of this dissertation, most of the known linguistic measures of entrenchment will be highlighted with respect to what these measures can tell us about entrenchment and why these measures alone cannot define entrenchment. Moreover, it will be highlighted that it is very likely that the interaction of several linguistic factors has a stronger effect on the speed and degree of entrenchment than any single factor alone. In essence, it is not enough if a word is very frequent and the transitional probability between two words is very high; the language user has to be familiar with that word or word combination in order to show any processing or production advantage. This claim sounds very self-explanatory; however, there are very few, if any, studies that analyse the interactions between several linguistic factors.

In addition to the driving forces, this dissertation assumes that entrenchment cannot happen without a range of cognitive skills that enable us to keep track of the driving forces of entrenchment. A word can be very frequent, but the language user has to register this frequency and keep track of it. In other words, a range of cognitive skills is necessary to transform information on the frequency of occurrence into a facilitatory factor of language processing and production. Entrenchment, interpreted from a psycholinguistic perspective, is the encoding of new memories, their storage and subsequent retrieval. Frequency and the other driving forces lead to the encoding of these words, phrases, and construction in mind; however, how easily and quickly these words become entrenched is subject to individual differences on the cognitive level. The second part of the theoretical chapter will highlight the cognitive factors that enable us to entrench linguistic units and emphasise the role of each skill in the process of entrenchment.

The final component of entrenchment, as viewed within the framework of this dissertation, is the end product: automaticity. Automaticity is a complex phenomenon on its own and refers, in a linguistic sense, to the effortless perception and production of different linguistic units and constructions. Automaticity can be observed in language and in many other domains, like learning how to type blindly or playing the piano. Thus, any repeated activity is stored in our memory and can be activated effortlessly and performed automatically. Automaticity in language can be manifested in reading, writing, and speaking fluency, effortless processing, and the accurate recall of linguistic items.

3 The driving forces of entrenchment

The entrenchment of linguistic units operates on the level of specific instances of language use in which a speaker takes part. Every time speakers process (i.e., interpret or produce) a linguistic structure, they add to earlier memories of linguistic events and strengthen the representation of that structure. The processing of linguistic structures presupposes that the speaker/listener not only tries to understand the meaning of the structure but also registers metalinguistic data such as whether they have already heard or seen the item, in which context they have observed it, and whether the item is usually preceded or followed by a particular item (e.g., collocations, idioms, phrasal verbs). The joint effect of the stored metalinguistic information determines how fast the processing of the item is. The more clues the language user has, the lower the item's activation threshold is. Frequency seems to be the most influential metalinguistic data that speakers register. It is also the most intensively researched factor related to entrenchment. However, recent empirical studies suggest that context- or probability-based measures of frequency might be better indicators of entrenchment. The rationale behind this suggestion is that words usually do not occur in isolation. Thus, the information provided by the context in which the word occurs should have a faciliatory effect on entrenchment. These forces will be discussed in the following section, beginning with the main factor, frequency and then moving onto contextual and predictability-based measures that are assumed to determine the speed and degree of entrenchment.

3.1 Frequency and entrenchment

Since the rise of usage-based theory, frequency has become one of the most studied factors in language acquisition, processing, and production. Frequency influences not only language use but also the mental representations that we have of the language and how we process these representations. The role of frequency in language comprehension and production was described before the emergence of the concept of entrenchment. Spoken and visual lexical access models treat frequency as the predictor that explains the most significant portion of variation in human performance. These models all share the assumption that word frequency is a kind of "counter in the head" (Baayen 2010: 439). Every repetition through auditory or visual stimuli or producing a word is supposed to increase this counter. However, it has to be highlighted that frequency, as much as it is discussed in the literature, is not a well-defined measure. There are various ways of quantifying frequency and various units upon which to base the count, e.g., single words, collocations, patterns, and constructions.

In order to avoid the circularity involved in the conceptualisation of entrenchment through frequency (the more frequent, the more entrenched; the more entrenched, the more frequent), the effects of frequency on the cognitive system and the language itself have to be separated. The most observable effects of frequency on language usage are the conserving and reducing effects (Bybee 2002b, 2007b; Divjak & Caldwell-Harris 2015). At first glance, these two effects seem to go in the opposite direction, but a closer look at the effects shows that they impact different paradigms of linguistic units. The conserving effect ensures that high-frequency words resist changes on the basis of analogy with other forms. For this reason, high-frequency irregular verbs do not undergo regularisation as the low-frequency irregular verbs tend to do (e.g., sleep / slept but weep/wept or weeped). Due to their tendency to occur very frequently in the language, irregular verbs have strong lexical representations, and these forms are unlikely to be replaced by new forms that were formed following regular patterns.

The reducing effect manifests that frequently used phrases and words are phonologically reduced. Phonological reduction is the result of the repetition of neuromotor sequences that consequently leads to a greater overlap and the eventual reduction of the component articulatory gestures (Bybee 2007a: 11) (cf. Kohler 1990). Just as the repetition of any behaviour increases the ease of performing that behaviour, the repeated pronunciation of words will lead to effortless and fluid production. However, the reducing effect can be explained by factors other than frequency. The psychological explanation emphasises that speakers always aim to produce linguistic expressions with minimal effort; at the same time, they try to ensure that what is said is clear to their interlocutor. Since frequent words are more expected (or predicted) by the listener, they are likely to be understood even when phonetically reduced.

Moreover, research has shown that the reduction depends on the context; thus, claiming that only frequency triggers reduction would be misleading. These two effects related to frequency show the opposite effect on low-frequency words. Infrequent items are resistant to phonological changes but are often subject to conceptually motivated changes[1].

[1] For an extensive review of the frequency effects on the language system see Behrens & Pfänder (2016b) and Bybee (2007b)

In addition to these two main effects, frequency has been shown to influence phonological variation (Coetzee & Kawahara 2013) and syllable frequency (Cholin et al. 2011). Apart from phonology, the effects of frequency are visible in morphology, especially in inflectional morphology (Bowden et al. 2010) and compound constructions (Baayen et al. 2010). In syntax, frequency influences the choice of tense/aspect (Ellis 2013), verb-argument constructions (Ellis et al. 2014), and different sentential complements (Ambridge et al. 2015).

Beyond the effects on language use, frequency causes changes in our mental representations. Numerous studies have shown that high-frequency words are (a) remembered and recalled more efficiently in both recall and recognition tasks, (b) more easily identified (Nusbaum & Dedina 1985), (c) less likely to be mispronounced and (d) correctly judged as high-frequency words in subjective frequency-estimation tasks (Ambridge et al. 2015)[2]

The way in which the manifestation of entrenchment and the frequency effects are reported in the literature implies that the two concepts are equal. Thus, frequency is a direct and quantifiable measure of entrenchment. This is the reason why some researchers (Stefanowitsch & Flach 2017; De Smet & Cuyckens 2007; Langacker 2008; Schmid 2000) suggest that defining entrenchment through the frequency of occurrence is the only logical way to approach the concept, and frequency should be regarded as a "quantitative signpost of the degree of entrenchment" (Mukherjee 2005: 225).

It is argued in this dissertation that entrenchment and frequency are related, but frequency alone does not define entrenchment; it only triggers the process through which entrenchment is achieved. The process that is triggered by frequency of occurrence or repetition is nothing else but learning and memory formation. Experience is the basis of learning, and different frequency measures are the ways in which we quantify our language experience. This separation of the driving forces from the process itself entails that frequency and entrenchment are discrete but inextricably interwoven. This approach gives the frequency of occurrence its rightful place as a fundamental shaper of the cognitive representations that are dynamically responsive to experience. Nevertheless, it emphasises that frequency alone cannot explain the speaker-dependent variation that determines which linguistic units are

[2] Comprehensive reviews on frequency and its effects on first and second language learning, representation and change are available in the work of Ellis (2002), Blumenthal-Dramé (2012), Divjak & Gries (2012).

entrenched and to what degree. Acknowledging the limitations of the frequency effect promotes the need to use corpora tailored to the participants of a given study; the choice of corpora should be motivated by the social and educational background of the participants. Moreover, it prompts the idea that a better understanding of the concept of entrenchment can be achieved if more linguistic factors (such as familiarity, contextual diversity, and transitional probability) are incorporated into the analysis because, in language comprehension and production, words are always embedded into more complex units and are not processed separately. The additional information gained from the context undeniably contributes to the entrenchment of the linguistic units. Therefore, instead of aiming to isolate the effect of the often-used linguistic factors related to entrenchment, it is more fruitful to treat them as factors that jointly influence the speed and degree of entrenchment. Examining how the interaction of multiple effects contributes to entrenchment can deepen our understanding of the whole concept.
To highlight the problems of analysing the linguistic factors separately, let us hypothesise how quickly and effortlessly a language user processes the bold-typed words in the following sentences.

(1) *People should not have to pay more for renewable* ***energy****.* (COCA, SPOK 2010 : 72)
(2) *She called the "epidemic" a "public* ***health*** *crisis".* (COCA, BLOG 2012 : 28)
(3) *The Washington Post reports on research that suggests your friends' social* ***media*** *posts are making you spend more money.* (COCA, SPOK 2019 : 43)
(4) *Russia has, shall we say, a problematic human* ***rights*** *record at a minimum.* (COCA, SPOK 2019 : 18)

These four sentences come from different academic fields and literary genres. A commonality among them is that all four bolded words are preceded by an adjective that, to different degrees, predicts the noun that follows. The main question is whether some of the boldface nouns show a processing advantage and whether the aforementioned linguistic factors explain it. The table below summarises the noun and bigram frequencies and forward and backward transitional probabilities of the four nouns and the adjectives that preceded them using three different corpora.

Table 1. Corpus-based measures of the selected nouns

Noun	Corpora	Noun frequency (pmw)	Adj+Noun Frequency (pmw)	Forward transitional probability	Backward transitional probability
	BNC	101,36	1,25	0,55	0,039
energy	COCA	164,65	4,78	0,65	0,9
	enTenTen18[3]	224,91	10,59	0,57	0,132
	BNC	159,57	4,25	0,01	0,09
health	COCA	300,08	17,63	0,06	0,17
	enTenTen18	288,64	14,16	0,05	0,02
	BNC	4,35	-	-	-
media	COCA	161,811	21,64	0,08	0,37
	enTenTen18	12,54	2,35	0,01	0,37
	BNC	903,03	15,67	0,11	0,21
right	COCA	1593,67	27,58	0,12	0,22
	enTenTen18	637,88	20,96	0,13	0,30

Using the four measures of entrenchment listed in the table, we could postulate the following hypothesis:

a) The word with the highest frequency (i.e., ***right,*** followed by ***health***) will show a processing advantage;
b) The word with the highest bigram frequency (i.e., ***media***) will show a processing advantage;
c) The word which is the most predictable from the context (i.e., ***energy***) will be processed the fastest;
d) The word with the highest backward transitional probability (i.e., ***right***) will be processed the fastest.

All these hypotheses are motivated by existing empirical research[4]. However, the question is whether it is possible to see any processing advantages among these words using only this linguistic metadata. Is the word ***right*** processed significantly faster than ***media*** just because it has a higher frequency? Empirical studies are available to validate the correctness of each hypothesis, which implies that, to a certain degree, all of these factors contribute to entrenchment to different degrees. Therefore, instead of looking for the strongest predictor of entrenchment, it is more beneficial to consider all the possible predictors and analyse how their combined effect contributes to the entrenchment of the target linguistic units. It is unlikely that speakers store only selected

[3] enTenTen18 is an English corpus made up of texts collected from the internet. It belongs to the TenTen corpus family. The corpus is available through Sketch Engine.

[4] For hypothesis **a** see Balota & Chumbley (1985); Balota et al. (2012)
For hypothesis **b** see Karinne Sauval & Chetail (2017)
For hypothesis **c** see Lorenz & Tizón-Couto (2019); Bell et al. (2009b)
For hypothesis **d** see McConnell & Blumenthal-Dramé (2019)

information about the occurrence of words; for instance, that they register frequency information but neglect the information provided by the context, such as the probability of the word, possible collocations, and register. This dissertation suggests that speakers take advantage of all available information during language production and processing. Therefore, language models that try to describe entrenchment should consider all possible metalinguistic data available to speakers.

Applying exemplar theory to the description of entrenchment would make it possible to incorporate several linguistic and social factors into the analysis of speakers' performance on different tasks. The core of exemplar theory is the idea that language acquisition is significantly facilitated by repeated exposure to concrete language input. Frequency of occurrence influences exemplar access for production, and recency of occurrence has an impact on both categorisation and production. As it emerges from this short description, exemplar theory operates on highly relevant premises for the description of entrenchment. Both exemplars and entrenched memories are in constant flux, with new input updating these memories and old unused memories fading away. Exemplars are complex units; they include information on the word's occurrence, pronunciation, syntax, morphology, and social factors, such as register and style. Therefore, thinking about entrenched linguistic units as exemplars would allow us to incorporate not only the linguistic metadata available to speakers but also social and cognitive factors that are inevitable for entrenchment to take place. This suggestion fits perfectly within the framework of the lexico-grammatical continuum introduced by the usage-based theory. Nevertheless, it has to be emphasised that even the combined effect of all these factors is not able to account for any individual differences, which could explain why a lawyer is likely to show a processing advantage for the combination ***human rights***, or an engineer for ***renewable energy***. To explain these individual differences among speakers, it is inevitable to extend the analysis of entrenchment by means of a thorough examination of the factors that focus on the social background of the speakers and the cognitive abilities that enable them to form memories.

3.2 Other predictors of entrenchment

As mentioned earlier, there are different ways of quantifying frequency and different units upon which we can base the counting. These different methods of counting frequency all aim to find the best way of quantifying our experience with language.

These additional measures of our language experience can be divided into two groups: corpus-based (such as contextual diversity, dispersion, transitional probability, and association measures), and those that aim to directly measure each language user's subjective experience with given words (such as familiarity judgement scores). Researchers are seeking new methods of quantifying language experience because word frequency count does not take into account the context in which words occur and how neighbouring words can influence language comprehension and production processes. During language comprehension, words usually do not occur in isolation. Sentence and word processing are likely to be influenced by predictions. These predictions and expectations are driven by our knowledge of typical patterns in language use. These patterns include typical collocations, constructions, and valency patterns, so when reading sentences like

(a) *To make solar cells a competitive alternative to other renewable* ... (EnglishWeb2020)

(b) *She insisted* … (EnglishWeb2020)

participants are very likely to anticipate seeing the word *energy* in the case of example (a) and the preposition *on* or the conjunction *that* in case of (b). High transitional probabilities between the words *renewable* and *energy* and between the verb *insist,* and the preposition *on* could be one of the possible explanations for these expectations. However, high-transitional probability is not the only possible explanation for why we anticipate seeing the noun *energy* or the preposition *on*. In the case of (a), the whole sentence is building up this expectation as all the other words are semantically priming the word *energy*. As far as (b) is concerned, the valency pattern of *insist* evokes some possible subcategorisations like *insist + that_CL*, *insist + on_NP*, etc. Moreover, it can be assumed that *renewable energy* and *insist on* are stored as chunks and are accessed in one single step. It is impossible to determine which explanation is true as they are all interrelated to a certain extent.

Some of the above-listed metrics (frequency, probability, contextual diversity) of language usage intensity have been predominantly used in theoretical linguistics, language acquisition and applied linguistics, while the others were mainly used to build computational models that aim to explain language production and comprehension. However, there is no consensus on which of these metrics are the most cognitively realistic since researchers have either pitted only some selected measures against each other or focused on one of them. To the best of my knowledge, no single empirical research project has incorporated all these measures into their analysis with the aim of analysing how their combinatory effect leads to entrenchment. Moreover, this dissertation aims to focus on the cumulative effect of these measures on language production and processing rather than finding which one is cognitively the most realistic. In the following subchapter, the psychological validity and the necessity to incorporate these measures into the analysis of entrenchment will be discussed.

3.2.1 Contextual diversity, dispersion, and semantic distinctiveness

The main aim of any corpus-based measure of frequency is to determine the importance of particular words or grammatical patterns for language teaching and to reflect the degree of their cognitive entrenchment. All three measures, contextual diversity, dispersion, and semantic distinctiveness, were introduced to linguistic analysis because empirical evidence suggests that they can better explain the language user's performance on different tasks than pure frequency. Nevertheless, it has to be noted that all three measures are calculated using the frequency of occurrence of words, i.e., they are not independent of frequency. In other words, these context-based metrics are highly influenced by how frequent the word is. These measures are calculated in the following ways[5]:

(a) counting the percentage of films in a subtitle corpus that contains the word (contextual diversity, according to Baayen (2010)).

(b) eliminating the double-counting of words that occur multiple times in the same document (contextual diversity, according to Adelman et al. (2006) and dispersion, according to Baayen (2010)).

(c) weighting each word count by the dissimilarity of the context the word appears in relative to each other, earlier occurrences (semantic distinctiveness, according to Johns et al. (2012)).

Among these three measures, contextual diversity is the most researched one. Empirical research suggests that contextually diverse words are acquired earlier in development and learned faster and more accurately (Rosa et al. 2017), and adults have a wide range of associations with these words (Hills et al. 2010). Moreover, experiments with young readers have shown that contextual diversity is the central determiner of word identification times (Perea et al. 2013), and large corpus-based studies have shown that contextual diversity predicts word-processing times

[5] For a critical review of dispersion and other adjusted frequency measure see Gries (2008)

independently of word frequency and that there is no evidence for a facilitatory effect of word frequency independent of contextual diversity (Adelman et al. 2006).
The main question here is why contextual diversity turned out to be a better predictor of language users' performance than word frequency alone. What is the extra bit of information that contextual diversity offers? It is already well established that repeated experience with or exposure to a particular word makes the word more identifiable. The rationale behind this assumption is that each and every exposure to a word has a long-term influence on its accessibility. In other words, frequent usage leads to entrenchment. However, research on memory has found that the extent to which the repetition of a particular item affects the later retrieval of that item depends not only on the mere number of repetitions but also on the separation of exposures in time and context (Glenberg 1976; 1979). Indeed, under some conditions, if neither time nor context changes substantially, there may be no benefit of repetition at all (Verkoeijen et al. 2004). If the human memory for word recognition operates under the same conditions, then it can be assumed that the effects of repetition will be diminished or eliminated when the repetition of the word always occurs in the same context. Word frequency and contextual diversity can be hard to separate sometimes; nevertheless, according to Vergara-Martínez et al. (2017), their effects on the human cognitive system might have a different origin. Word frequency facilitates lexical access, and contextual diversity leads to a semantically richer representation. Referring back to exemplar theory, every occurrence of the word strengthens the exemplar's position in the mental lexicon; the diverse contexts in which the word occurs widen the scope of the exemplar with information about the text types where it is frequently used, its collocations and grammatical patterns. Therefore, word frequency and contextual diversity should be seen as complementary and not competitive measures. It is undeniable that words experienced in different contexts are easily recognised; simultaneously, occurrence in different contexts adds to the general frequency of the word.
The only difference between contextual diversity as defined by Adelman et al. (2006) and Baayen (2010) is that Baayen (2010) uses subtitle corpora to determine the word's contextual diversity, and Adelman et al. (2006) use general corpora. Empirical research has shown that word frequency counts obtained from subtitle corpora improve reaction times' predictivity compared to standard text-based frequency counts. Despite the scripted nature of subtitles, these corpora seem to resemble everyday language use better (see Brysbaert & New (2009) for further arguments in

favour of using subtitle corpora). Subtitle corpora include more short, simple words with strong emotional connotations. These factors facilitate rapid visual uptake of the information. This rapid visual uptake is undoubtedly required for lexical decision and word naming tasks where words are presented in isolation without any context. Baayen et al. (2016: 1130) argue that this indicates that frequencies taken from subtitle corpora provide excellent fits for experimental studies not because they capture the frequency information that drives participants' responses in them but rather because, as a register, subtitles serve to strongly confound frequency with several other variables that also contribute to faster lexical responses. However, one has to keep in mind that subtitle corpus-based frequency is a good predictor of human performance when measuring lexical decisions and word naming but loses its facilitatory power in other tasks. It is very likely that language users easily recognise words that they frequently hear in movies or television series, but it does not mean that they use these words in their everyday lives. Production experiments usually focus on complex grammatical patterns or specialised vocabulary, which are less likely to be found in subtitle corpora. Jones et al. (2012) even further developed contextual diversity by taking into account how semantically dissimilar from one another the documents in which the word appears are. The more dissimilar the texts are, the higher the word's semantic distinctiveness. Empirical studies (like by Johns et al. (2012) validated the claim that semantic distinctiveness has higher explanatory power than pure word frequency count or even contextual diversity.

It seems that there is an emerging tendency to come up with newer and more sophisticated metrics that are claimed to explain human performance. This tendency might seem like a competition between researchers who can come up with a measure that would explain the highest variance in a statistical model. The more sophisticated a measure is, the more artificial it appears. Because of the complexity of the cognitive system and the way language is processed and produced, it is very unlikely that one single measurement or factor would be enough to describe the underlying processes. It is more beneficial to explore the interaction of linguistic, social, and cognitive factors than to seek out the most powerful measurement to explain the observed variation in the language users' performance. In the following section, two more linguistic measures will be described that are likely to contribute to entrenchment and explain language users' performance on different linguistic tasks.

3.2.2 Predictions and probabilities

Transitional probability is an additional measure discussed in connection with entrenchment, which aims to explain why certain words or word combinations are processed faster and more accurately. It takes the linear character of language as the basis and builds upon its predictive nature. Observing words in predictable contexts can facilitate the process of entrenchment because it can lead to the automatisation of a sequence of separate items. Once a member of that sequence is encountered, that leads to automatic activation the next time. A further explanation is that the frequently co-occurring items undergo chunking, and these chunks are stored and activated holistically (Lorenz & Tizón-Couto 2019). This implies that predictable words are entrenched easier because the context itself leads to preactivation of the item that consequently strengthens its mental representation. Transitional probability can be seen as one possible way to quantify the predictive nature of langue together with cloze probabilities and entropy. Transitional probabilities, as defined by Jurafsky et al. (2001a), refer to the likelihood of one word occurring given the previous word. It is calculated using the observed frequencies of the bigram and the predicting word.

Prediction in language processing is a highly debatable question. Some researchers argue that predicting upcoming information ahead of time would be an unnecessary waste of processing resources (Jackendoff 2002; van Petten & Luka 2012) and that processing words when they occur may be more efficient and economical than predicting them in advance (Huettig & Mani 2016). The opponents of this view suggest that given the noisiness, ambiguity and speed of our linguistic input, prediction is the most efficient solution for fast, efficient and accurate comprehension (Kleinschmidt & Jaeger 2015; Kuperberg & Jaeger 2015), and prediction in the form of transitional probability can explain the phonological variation (especially temporal and spectral reduction) in language production (Jurafsky et al. 2001a; Bell et al. 2009b). The rationale behind assigning such a prominent role to predictions in language processing is that we build up certain expectations about upcoming words based on a range of linguistic and extralinguistic factors. In other words, we store the typical collocation and grammar patterns of the words, and additionally, we use our experience to predict what is a plausible continuation of the sentence. Empirical research has shown that probabilistic information about words, phrases, and other linguistic structures is represented in the minds of language users and plays a role in language comprehension (Jurafsky 1996; MacDonald 1993; McRae et al. 1998;

Narayanan & Jurafsky 1998; Trueswell & Tanenhaus 1994; Lorenz & Tizón-Couto 2019; McConnell & Blumenthal-Dramé 2019), production (Gregory et al. 1999; Roland & Jurafsky 2000; Gahl & Garnsey 2004; Warner & Tucker 2011), and learning (Brent & Cartwright. 1996; Saffran et al. 1996; Seidenberg & MacDonald 1999). Children as young as eight months old are able to calculate transitional statistics about the frequency of syllable occurrences and use these statistics to segment continuous speech streams without explicit acoustic cues as to the boundaries between words in the input (Saffran et al. 1996). Moreover, eye-tracking studies have revealed that readers focus less on predictable than unpredictable words (Balota et al. 1985; Ehrlich & Rayner 1981; Rayner et al. 2001; Boston et al. 2008; Demberg & Keller 2008; Frank & Bod 2011; McDonald & Shillcock 2003; Smith & Levy 2013; Staub 2015). Smith & Levy (2013) go as far as to claim that prediction is necessary and essential for every task the brain performs; thus, it is a fundamental principle underlying its operation[6].

These empirical results make a strong argument for including some prediction measures in the analysis of the language users' performance of different language tasks in particular and in the theory of entrenchment in general. Nevertheless, some unanswered questions suggest that prediction alone cannot explain the fluency we observe in language processing and production. Among these unanswered questions are: (a) what is the function of prediction in language processing? (b) what are the cues used to predict upcoming words? (c) what mechanisms are involved in predictive language processing? (d) do individuals always predict upcoming words? (Huettig 2015).

Based on numerous studies, the most apparent answer to question (a) is that we need prediction to acquire new things. Indeed, some psycholinguists argue that learning happens only when prediction happens (Chang et al. 2013). Huettig (2015) claims that our capacity to extract forward statistical regularities is not a straightforward indication that this process is driven by prediction. Empirical evidence shows that learning can happen without any predictive processes (Pelucchi et al. 2009a, 2009b; Perruchet & Desaulty 2008).

The regularities found in language are the cues we use to predict upcoming words. The availability of these regularities in language does not automatically mean that language users always rely on them and that there are no individual differences

[6] An ERP study by van Berkum et al. (2005) revealed that people can indeed predict upcoming words in fluent discourse and, moreover, that these predicted words can immediately begin to participate in incremental parsing operations.

among language users concerning the degree to which they rely on these regularities. The fact that collocations or typical grammatical patterns are recalled and used effortlessly can also be seen as their encoding in long-term memory. To experimentally validate which assumption is valid is almost impossible. However, we do not rely only on linguistic knowledge when we predict upcoming words; our extralinguistic and world knowledge helps determine what is plausible in a particular situation. Thus, when seeing a sentence like *The sky is covered with dark clouds, so do not forget to take your…*, one would know without using any kind of predictive process that the word *umbrella* is meant. The reader or listener of the sentence knows what a likely continuation of the sentence is because they have a concept about a thing that protects us from the rain. For this thing, English has the name *umbrella*. There are, of course, other situations when language users rely on predictive processes. In other words, prediction is gradual, just like all other phenomena that make language comprehension and processing easier.

Turning to the mechanisms that underlie the predictive process, several different solutions have been proposed by researchers. In recent years, the dual system approach has gained much support (Huettig 2015). It claims that cognition operates on two systems; one of them is responsible for the automatic processes performed unconsciously and without much effort. The second system is responsible for directing attention towards effortful activities. This division between the systems is frequently discussed in relation to language learning (Ellis 2015) and the division of our memory system (Llompart & Dąbrowska 2020). The dual system model can easily accommodate the gradual nature of predictivity. The retrieval of entrenched items from memory is likely to rely on the predictive process, while those items that are less frequently observed require conscious processing and effortful production. Moreover, this system allows taking into account the individual differences among language users. The significance of factors such as working memory or general cognitive efficiency is less researched. However, it cannot be denied that these cognitive factors greatly influence how we elicit information from the given input and how sensitive we are to the frequency and context-based cues available during language processing and production.

One of the emerging issues that the literature review reveals is that the existing research shows an emerging pattern where prediction is treated in an all-or-nothing manner. Language processing and acquisition cannot exist without prediction or the other end of the continuum – prediction is entirely unnecessary for these processes.

One of the possible ways to approach the problem is to assume that prediction alone might not be powerful enough to explain language processing and acquisition, but, without a doubt, it can be regarded as a "helping hand" that, in certain situations, fosters not only language acquisition but also processing and production. This approach fits the conclusions offered in the previous section on the role of frequency-based metrics. The human cognitive system is too complex to rely on one single measure. Language users get the most out of the available input and use it later to ease processing and production. In other words, it is suggested in this dissertation that the more linguistic measures are incorporated into the analysis, the more we can learn about the underlying processes that enable us to comprehend and produce language so effortlessly. It is likely that for processing one part of the sentence, the listener relies more on transitional probabilities, e.g., processing words within a phrase boundary, while outside of the phrase boundary, they might rely on the context or on frequency. Thus, in this dissertation, it is postulated that the combinatory effect of frequency and corpus-based measures and the individuals' general cognitive abilities are likely to explain the variation we observe in language production and processing, and no single factor or metric has the power to do it without considering its interaction with other predictors.

3.2.3 Familiarity

Familiarity is the only measure in the list of metrics that takes into account the individual differences in language users' experiences. The extent to which a linguistic unit is entrenched varies from person to person and over time. Familiarity scores can indicate these differences in the degree of entrenchment. Familiarity as a concept rests upon two factors: a) the frequency of the linguistic unit, construction or specific phrase, and b) the similarity to existing linguistic units, constructions or specific phrases (Bybee 2010: 214). Thus, familiarity "taps into exposure and chunking" (Verhagen 2020: 11). It reflects how frequently the speaker has used the linguistic unit and how often they have observed it.

Moreover, familiarity scores can be viewed as a simple tool for measuring the extent and type of previous experience language users have had with each linguistic item. In other words, variation in the familiarity scores can indicate differences in the degree of entrenchment of the specific words. Furthermore, these scores show to what extent the degree of entrenchment is part of the speaker's knowledge, i.e., whether the linguistic units that were given higher familiarity scores were also processed faster.

Familiarity scores can be obtained from large databases such as MRC (2006)or Wilson's List (1988) or directly from familiarity judgement tests taken by each experimental participant. These databases offer a convenient way to gather information that reflects the general tendencies observed in language; however, they are unlikely to be fully representative of the linguistic experiences of the people taking part in a study and unlikely to be equally representative for all participants alike. Typically, judgments by different participants in those large databases are averaged, and inter-individual differences are regarded as 'noise' (Verhagen 2020). People, most likely, differ to a larger extent in how and how often they encounter and use particular combinations of words and chunks. The frequency with which users encounter the words may depend on their social background, and the ease with which they arrive at certain abstractions is most likely to depend on their cognitive capacities. For example, the word *peer* is not frequently used in everyday language. However, there are some fixed collocations where the word appears frequently, e.g., *peer pressure, peer-reviewed journal,* or *peer support*. Speakers in the field of academia might be exposed more frequently to these multiword units, which is expected to be reflected in their familiarity scores. Moreover, language users would likely rate the word differently when given in isolation or in a typical context. Thus, differences in familiarity scores can shed new light on how words are represented in our mental lexicon: as single words, chunks, or exemplars.

A substantial number of studies have made use of familiarity ratings for words, word pairs, phrases, idioms, and metaphors. These ratings were found to be significant predictors of reading times (Cronk et al. 1993; Juhasz & Rayner 2003; Williams & Morris 2004), as well as performance on lexical decision and speeded naming tasks (Gernsbacher 1984; Connine et al. 1990; Blasko & Connine 1993; Juhasz et al. 2015), speeded semantic judgment tasks (Tabossi et al. 2009), and perceptual identification tasks (Caldwell-Harris et al. 2012).

One might ask why familiarity judgements outperform the significance of corpus-based frequency. It is expected to find a strong correlation between frequency and familiarity. Tanaka-Ishii & Terada's (2011) analysis of the correlation between word frequency and familiarity suggests that for a given corpus, high frequency is a necessary condition for a word to attain a high familiarity rating but is not a sufficient condition to make the word familiar. Thus, high frequency always indicates high familiarity but not the other way around, as certain groups of speakers are highly familiar with words that are particular to their field of work. Due to this same reason, familiarity can be considered a better predictor of human performance on certain tasks where specific genres and registers are investigated.

Both corpus-based frequency measures and familiarity ratings from large databases lack the possibility to analyse individual differences. Verhagen (2020) extended this line of research by investigating intra- and inter-individual differences and the reliability of metalinguistic judgments in the form of familiarity judgements. With the help of numerous experiments, Verhagen (2020) empirically verified the reliability of familiarity judgement tasks and proved once again that familiarity ratings are valuable metalinguistic information. The magnitude estimation ratings did not differ significantly from the ratings expressed on the Likert scale, as the experiments' results have shown. She concluded that a Likert scale allows examining whether the stimuli on average were more familiar to the participants and that in Likert ratings, variation across items and times were more pronounced in low-frequency items than in high-frequency ones.

As Taylor (2012: 250) puts it: "It is evident even to the most casual observer that speakers of the 'same' language may exhibit variation in their usage patterns according to their geographical provenance, their social status, their educational background, their age, gender, ethnicity, and so on". What was in the earlier research considered noise contains essential information necessary to understand better the processes underlying entrenchment and the role individual differences play in this process.

3.2.4 Conclusions

All things considered, it can be said that entrenchment is a complex phenomenon that cannot be explained by relying on single factors. Research so far has focused on comparing different linguistic factors with the aim of finding the one that explains the most variation in the speakers' performance. As the literature overview reveals, there is plenty of research validating the significance of each corpus- or context-based metric for the process of entrenchment. This fact alone suggests that all these metrics, to a certain extent, are necessary for the process to take place. Just as language users extract all the available cues from the input, researchers are encouraged not to rely on the significance of single factors. It is very unlikely that our cognitive system uses only frequency information or only prediction when processing or producing language. Consequently, incorporating a range of linguistic factors into the analysis allows us to analyse how the cognitive system gets the most out of the input.

4 Entrenchment as a process

Within the framework of this dissertation, entrenchment is defined as a phenomenon that stands on three pillars; the cognitive processes that are responsible for quantitative and qualitative changes in the mental lexicon make up one of these pillars. These changes comprise the encoding, storage and strengthening of frequently observed and used linguistic structures. In other words, the linguistic factors mentioned in the previous chapter trigger certain processes that will result in the automatic processing and production of entrenched items through their encoding into long-term memory. This automaticity is the end stage of entrenchment. The processes responsible for the quantitative and qualitative changes in the mental lexicon include Hebbian, statistical and surprised-based learning. The above-listed learning processes contribute to memory consolidation, the formation of holistic units, so-called chunks, schema formation, and routinisation. In the same way, as numerous linguistic factors trigger entrenchment, several processes contribute to entrenchment. These different processes are likely to contribute to the entrenchment of different structures to different extents, e.g., statistical learning is likely to help us entrench words and collocations. At the same time, schematisation could be a more dominant process for the entrenchment of constructions and more schematic language patterns.

However obvious it sounds, the processes mentioned above are taking place in individual speakers' minds. Accordingly, they are more or less subject to individual, speaker-related differences by definition. Most of these differences are difficult to grasp and empirically investigate as their sources are hidden in the language users' prior exposure to the language and their cognitive abilities. This section of the chapter attempts to elucidate how these individual differences influence the ease and degree of entrenchment.

From a neurobiological and psychological point of view, entrenchment is essentially an experience-based learning process. The more frequently a language user has been exposed to the linguistic item, the more strongly this item will be represented in their mind. The first part of the chapter will discuss the three different learning concepts, followed by schematisation, routinisation, chunking, and memory consolidation theories.

4.1 Learning

All kinds of learning, including language learning, crucially depend on input and how often and in which context this input is observed or produced, i.e., how often it is repeated. One of the most widely studied characteristics of human memory is that repeating the same material yields better memory for that material. This is the rationale behind devoting so much attention to the different frequency measures when discussing entrenchment. Three existing models that are relevant for entrenchment and are based on repetition are enlisted by Schmid (2020). The three learning models are Hebbian learning, statistical learning, and surprise-based learning. The discussion of these models will start with Hebbian learning.

4.1.1 Hebbian learning

In 1949, the great Canadian psychologist Donald Hebb produced one of the first theories that analysed the biological basis of learning (Baddeley et al. 2020). Hebb proposed that long-term learning is based on cell assemblies. A cell assembly arises when two or more nerve cells fire simultaneously. During the simultaneous firing of the nerve cells, the synapse (i.e., the gap between two separate neurons) is constantly decreasing. This leads to the strengthening of the connection between the neurons. This process is often summarised by the phrase "neurons that fire together wire together" (Baddeley et al. 2020). Hebb proposed that long-term learning is the development and growth of synaptic connections. This particular learning process is defined as Hebbian learning. Accordingly, the core concept of Hebbian learning is based on repetition. The previous chapter discussed the vital role frequency of exposure and usage intensity play in the process of entrenchment. Thus, it can be assumed that Hebbian learning is a way to achieve the entrenchment of linguistic structures[7], rules, and items. This synaptic connection can represent the sounds in a word that are connected together, multi-word units or even abstract constructions. Schmid (2020: 207) emphasises that to understand this model fully, one must add the concept of anti-Hebbian learning to the picture. The catchphrase used to describe this phenomenon is “neurons out of sync delink”. In essence, if the connection between the neurons is not activated for a longer time, the memory representation might get weaker or completely lost. This process is part of Langacker's (1987: 59)

[7] A detailed description of the role of Hebbian learning for language acquisition is provided by Garagnani et al. (2008)

definition of entrenchment: "Every use of a [linguistic] structure has a positive impact on its degree of entrenchment, **whereas extended periods of disuse have a negative impact**" (emphasis added by the author). This is the reason why it is cognitively demanding to recall words that we rarely use.

As mentioned above, according to the Hebbian learning model, cell assemblies are assumed to form the neurobiological representations of cognitive elements, such as gestalt-like figures or words. These neurobiological representations are entrenched memories. As noted by Pulvermüller (1999), the representations of words are to be found in well-defined areas of the cortex. Neurons that are part of the assembly due to frequent co-activation will turn into functional units. The same can happen to multiword units; their frequent coactivation will strengthen the synapse between word A and word B so that later word A influences the activation of word B and vice versa. This final claim is based on holistic storage and chunking. This process is described by Pulvermüller (1999: 255) in the following way: "If a sufficiently large number of the assembly neurons are stimulated by external input (either through sensory fibres or through cortico-cortical fibres), activity will spread to additional assembly members and, finally, the entire assembly will be active." Due to this chain reaction type of activation, certain empirical studies (Divjak & Caldwell-Harris 2015) have found that transitional probability is a better predictor of entrenchment than pure frequency, as this measure takes into account the likelihood of one word being proceeded by another one (for more details, see chapter 3.2.2). However, it cannot be surely stated that Hebbian learning is driven either by frequency, transitional probability, or other predictors. It is very likely that they all contribute to the formation of synaptic connections to different degrees.

To investigate this connection between neurons is much easier using artificial intelligence and computational-based models. These models interpret learning as the weight changes in the connections between the nodes (Schmid 2020). These nodes have different activation thresholds. The threshold can change depending on the number of inputs that the node previously has received. An additional factor that influences the threshold is the weight of the input. In human communication, this weight can be understood as salience, and the amount of input can be any kind of frequency measure both in production and perception. According to Ellis (2013: 7), salience is the general perceived strength of stimuli. Salient items and features are attended to and more likely to be perceived and enter into subsequent cognitive processing and learning. Expectation-driven processes can drive attention that is triggered by salient items. In essence, not only does the frequency of occurrence matter but also how important the input is for the speaker. In communication, high-frequency words and patterns are processed much faster compared to low-frequency words and require less cognitive effort as they have lower activation thresholds. The different linguistic factors described at the beginning of the chapter are expected to jointly influence this threshold level.

In a nutshell, the Hebbian learning model tries to accumulate the role frequency, predictability, attention, and context play in learning and the subsequent formation of memories of newly acquired words, patterns, or skills. In the end, it is a model that tries to understand how our brain learns, and despite the development of empirical methods, it is still very difficult to exactly describe the learning process, let alone fully understand how the brain produces and comprehends language.

4.1.2 Statistical learning

From the psycholinguistic perspective, entrenchment as a process is best described as a form of implicit statistical learning (Frost et al. 2015; Jost & Morten 2017; Reber 2013). To date, empirical research has found evidence for statistical learning in different domains, including segmentation of auditory input (Pelucchi et al. 2009b; Warker et al. 2008), visual search (Baker et al. 2004; Fiser & Aslin 2002), contextual cueing (Goujon & Fagot 2013), and visuomotor learning (Hunt & Aslin 2001).

Statistical learning can be defined as a domain-general process by which humans implicitly form associations between stimuli by tracking and storing the statistical relationships between them (Schmid 2020)[8]. The obtained information can be used to generate predictions about upcoming words and structures. A slightly different view on learning in general and statistical learning, in particular, is expressed by Frost et al. (2015), who define learning as the process responsible for updating internal representations based on the received input. They understand statistical learning to reflect updates based on "the discovery of systematic regularities embedded in the input, and [to] provide a mechanistic account of how distributional properties are picked up across domains, eventually shaping behaviour" (Frost et al. 2015: 118). They suggest that statistical learning involves a set of domain-general neurobiological mechanisms for learning, representation, and processing that detect and encode a

[8] Research by Bogaerts et al. (2020) foster a bilateral integration of SL research with cognitive science: not only should domain-relevant evidence about the complexity of real-world input become more tightly integrated into SL research, but non-SL studies should also carefully consider the nature and range of statistical regularities that may affect learning and processing in a given domain.

range of distributional properties within different modalities or input types. These neurobiological mechanisms are frequently discussed in relation to entrenchment, which signals that these two processes are related.

Statistical learning is not a unitary learning system; rather, it consists of several neural networks operating in different cortical areas. This implies that the encoding of different input types varies depending on the neural network that is responsible for the encoding; thus, they are modality-specific. In simple words, statistical learning over auditory input differs from the process that occurs during statistical learning triggered by visual input. It is due to the fact that the auditory cortex is sensitive to temporal information and the visual cortex to spatial information. These inherent differences are reflected in how the sensory input is eventually encoded into internal representations (Frost et al. 2015). Figure 1 below represents the schematic account of this approach developed by Frost and colleagues (2015). It shows how the same learning and representation principles result in modality and stimulus specificity because they are instantiated in different brain regions, each with its characteristic constraints.

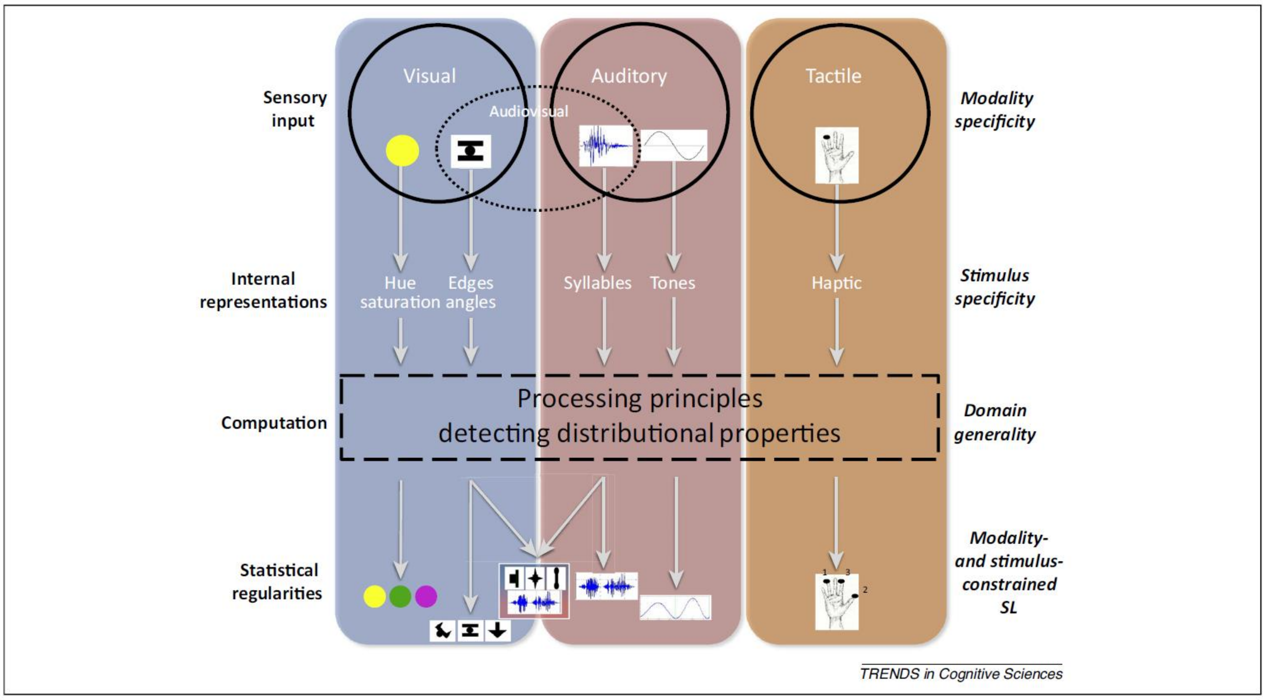


Figure 1. A theoretical model of statistical learning. Schematic representation of distributional information processing in the visual, auditory, and somatosensory cortex (Frost et al. 2015: 119)

Hebbian and statistical learning can also be viewed as not completely separate but complementary processes. Hebbian learning is the first step to creating a mental representation of the item. If these mental representations form a pattern in the input, statistical learning helps detect and register it. Both of these learning processes can lead to the formation of chunks but differently. Hebbian learning fosters the simultaneous activation of multiple items, and statistical learning is generally defined as the cognitive ability that allows us to form chunks (Jost & Morten 2017: 235). Language users' sensitivity to adjacent linguistic units allows for syllables to become associated with one another to form words and for words to become associated with one another to create chunks. Learners track relationships between linguistic elements over the course of their experience and use the information in these relationships to continuously update their expectations and representations: statistical learning abilities can be thought of as mediating the effect of linguistic experience (Jost & Morten 2017: 237). This continuous update of their expectations and representations leads to the formation of chunks that later can be retrieved as a holistic unit. Thus, if language users have registered these relationships between the words, it can lead to the simultaneous activation of the neurons that represent the linguistic items. In other words, Hebbian learning describes the formation of chunks on the neural level and statistical learning on the psychological level. It is very likely that among the linguistic factors that are considered the driving forces of entrenchment, high transitional probability between words fosters statistical learning to the greatest extent.

Statistical learning is an inborn cognitive mechanism; nevertheless, recent studies point towards individual variation in statistical abilities. In several areas of cognitive science, it is now well established that understanding the source of individual differences holds the promise of revealing critical insight regarding the mental operations underlying performance, leading to more refined theories of behaviour. Empirical evidence (Reber 2013; Reber & Allen 2013) shows that individual differences in implicit learning skills are already visible in preverbal infants. Their learning skills greatly determine language skills later in life. Additionally, studies of individual differences conducted with adult speakers reveal that statistical learning scores were an even better predictor of their language skills than verbal working memory tasks (Misyak & Christiansen 2012). The study by Misyak & Christiansen (2012) has shown that individual differences in detecting adjacent and nonadjacent dependencies may map onto variations in language users' ability to track long-distance relationships spanning across lexical constituents (e.g., relating the object of an embedded clause to the subject and main verb of the sentence).

There are two primary sources for individual differences in one's statistical learning capacities (Frost et al. (2015)):

a) the variance related to efficiency in encoding representations within modality in the visual, auditory, and somatosensory cortex;

b) the variance associated with the relative computational efficiency of processing multiple temporally and spatially encoded representations and detecting their distributional properties.

Besides identifying the sources of differences, Frost and colleagues proposed empirical methods to test these differences[9].

Schmid (2020) suggests that statistical learning over syntagmatic associations[10] relates to the classic conception of entrenchment as unit formation. Language users use their statistical learning skills to update their knowledge of lexical and grammatical co-occurrence tendencies regularly.

Previous studies have shown that there is huge variation across individuals in their abilities to pick up regularities, given their linguistic experience. These differences highlight the importance of statistical learning in the process of entrenchment; increased experience with specific structures leads to more automatic processing and more fluent production. Through this constant updating, speakers become aware of collocations and other co-occurrence phenomena, allowing them to generate predictions of upcoming forms and meanings. In essence, statistical learning skills enable language users to track regularities in language; the brain fixes this regularity in the form of neural circuits, and the constant simultaneous activation of the neural circuits leads to the rise of multiword chunks that are holistically stored and can be retrieved from memory in one single step. There is robust evidence to support the importance of statistical learning for language acquisition and usage in general and entrenchment in particular. Numerous studies have found that between one-third and one-half of the language we use consists of fixed multiword stretches of speech (Conklin & Schmitt 2012). Without the ability to track systematic regularities embedded in the input, it would not be possible to have such a large percentage of formulaic sequences in the language. As a result of statistical learning, humans can produce and process language, not in a word-by-word manner but in larger chunks, making language perception and production much smoother and more fluent.

[9] For an extensive description of the methods see Frost et al. (2015)

[10] Schmid defines entrenchment as a process operation over four different types of associations: symbolic, syntagmatic, paradigmatic, and pragmatic. For an elaborated discussion see Schmid (2017a); Schmid (2020)

4.1.3 Surprise-based learning

The last type of learning process is founded on the claim that unexpected, surprising input leads to faster and more successful learning. Surprisal is defined by Barto et al. (2013) as a result of the discrepancy between what is expected and what is observed/experienced. The effect of surprisal in the learning process was established back in the 20th century by Rescorla & Wagner (1972: 75):

> The central notion here can also be phased in somewhat more cognitive terms. One version might read: organisms only learn when events violate their expectations. Certain expectations are built up about the events following a stimulus complex; expectations initiated by that complex and its component stimuli are then only modified when consequent events disagree with the composite expectation.

The composite expectations mentioned in the definition are the weighted sum of the saliencies of the stimuli in the constellation, according to Barto et al. (2013)[11].

Previous research has empirically validated that surprise is of great value for learning. Surprising input provokes more intensive processing of the to-be-learned material (Adler 2008). Moreover, it can increase the retention of information (Munnich et al. 2007). Upon encountering a surprising or unexpected input, people try to find an explanation for the observed input, which produces richer memory encoding (Foster & Keane 2019). This surprisal effect is furthermore observed in the event-related potentials (ERP) in the brain. The two most frequently investigated ERPs are the N400 and the P600 (Kutas & Federmeier 2011). Both of these ERPs show that when the input contains something unexpected, it triggers various reactions in our brains. This unexpected input attracts our attention and triggers memory encoding, leading to the faster entrenchment of the linguistic item. This type of learning is even more successful when the input evokes emotions from the learner (Divjak & Caldwell-Harris 2015). Schmid (2020) claims that statistical learning is, in fact, not a typical form of learning as there is no change in knowledge or behaviour during the process of strengthening the associations between grammar patterns and words or between words. He explains this claim by viewing it from an information-theoretical point of view. From this angle, "another encounter of a perfectly familiar piece of experience has zero entropy, i.e. no gain in information" (Schmid 2020: 209). In essence, repeated exposure to an item is less likely to change its behaviour due to its excessive

[11] Salience was already mentioned in the discussion of Hebbian learning. It implies that salience plays a crucial role not only in the process of entrenchment but also in learning any kind of skills in general. Günther et al. (2017: 289) define salience as the conspicuity of a stimulus relative to its surrounding items. Schmid (2020) argues that salience, just like frequency, is a very good example of how a self-reinforcing feedback loop interweaves usage and entrenchment.

entrenchment. A change of behaviour, i.e. learning, can occur, in a narrow sense, when language users "break their routine because they are surprised" (Schmid 2020: 209). Contextual diversity, dispersion, semantic distinctiveness, and entropy are all measures that try to quantify this surprisal effect. However, it has to be stressed that the effect of surprisal cannot occur unless language users rely on some predictive processes. In other words, if language users were not expecting certain words to occur in the input, they would not be surprised when a word is heard or observed in the input. Furthermore, surprisal is a function of the input's conditional probability given the preceding context, corresponding to how predictable the input is, and has been shown to influence processing costs as well as production choices (Zarcone et al. 2016: 2). Consequently, these different learning processes are interwoven as the linguistic driving forces behind them overlap. Thus, separating their effect on memory formation in particular and entrenchment in general, seems rather difficult and disadvantageous.

As mentioned earlier, the different linguistic structures (e.g., words, collocations, valency construction patterns) are likely to be learned by applying different learning processes. Language is rich in different constructions and valency patterns; some are more lexicalised, while others are more schematic. We need different processes and linguistic factors to entrench all these types successfully. Moreover, along this line of reasoning, the particular learning process employed by learners on particular occasions may vary between different individuals (Street & Dąbrowska 2010; 2014; Dąbrowska 2012, 2015a).

Surprise-based learning is thus based on the bottom-up mechanism and allows language users to learn new lexical items, collocations, constructions, and valency patterns. New and inexperienced items are likely to affect the items' representations to a higher degree than the rehearsal of already entrenched items (Schmid 2020; Jaeger & Weatherholtz 2016; Kleinschmidt & Jaeger 2015). Moreover, as mentioned earlier, a single exposure to the item might result in its learning and entrenchment without further repetition if it is very salient for the speaker/listener in that communicative situation (Divjak & Caldwell-Harris 2015).

4.2 Memory consolidation, chunking, and schematisation

After the words, collocations, patterns, constructions, and other linguistic structures have been extracted from the input with the help of statistical or surprise-based learning skills, they have to be stored in our memory. Given that the process of entrenchment entails that cognitive processes transform input into linguistic knowledge, entrenchment as a multifaceted concept should take into account the effects of entrenchment on the creation and updating of memories (Schmid 2020). Human memory has an impressive capacity to process information; however, its capacity to simultaneously activate items is strongly limited (Günther et al. 2017). Chunks help speakers organise and compress information into units and easily retrieve them when needed. In essence, chunks can increase this limited capacity since prefabs, formulaic expressions, and collocations are activated as one unit instead of separate words and patterns. This consequently requires less cognitive effort and time. This section will discuss three issues related to entrenchment as a process. First, memory consolidation will be discussed, followed by the role of chunking for memory and, accordingly, for entrenchment, and finally, the processes of routinization and schematisation will be discussed as certain forms of information compression.

4.2.1 Memory consolidation

Usage-based approaches treat language as an emergent product of our prior linguistic experiences; consequently, our knowledge should be described as a complex system of memories. Thus, everything we know, including words, patterns, rules, and prefabricated units, are stored in our memory, and during reading, writing, or listening, we activate these stored memories. Consequently, the encoding and updating of these memories is a vital part of the process of entrenchment. Every item that is entrenched has a memory representation; the degree of entrenchment depends on the strength of the memory representation. The linguistic forces discussed at the beginning of the chapter contribute to the degree of entrenchment. The learning strategies lead to memory formation and constant updating. While surprise-based learning can theoretically be more relevant for the formation of new memory representations, Hebbian learning is likely to influence the strength of mental representation.

Divjak (2019) defines memory as our ability to transform learned materials and skills into persistent knowledge. The memory system that holds this acquired information and knowledge needs to be flexible to incorporate incoming information and update the previously stored representation. Moreover, the memories that we have stored should stay stable and resist being forgotten despite the continuous flow of information and updating. As discussed in the framework of Hebbian learning, memories are assemblies of neurons formed during the simultaneous activation of those neurons. Memory consolidation is described as the process by which a temporary, labile memory is transformed into a more stable, long-lasting form (Squire et al. 2015). This consolidation is facilitated by repetition. In order to understand the above-described process, the memory system with all its component parts has to be discussed.

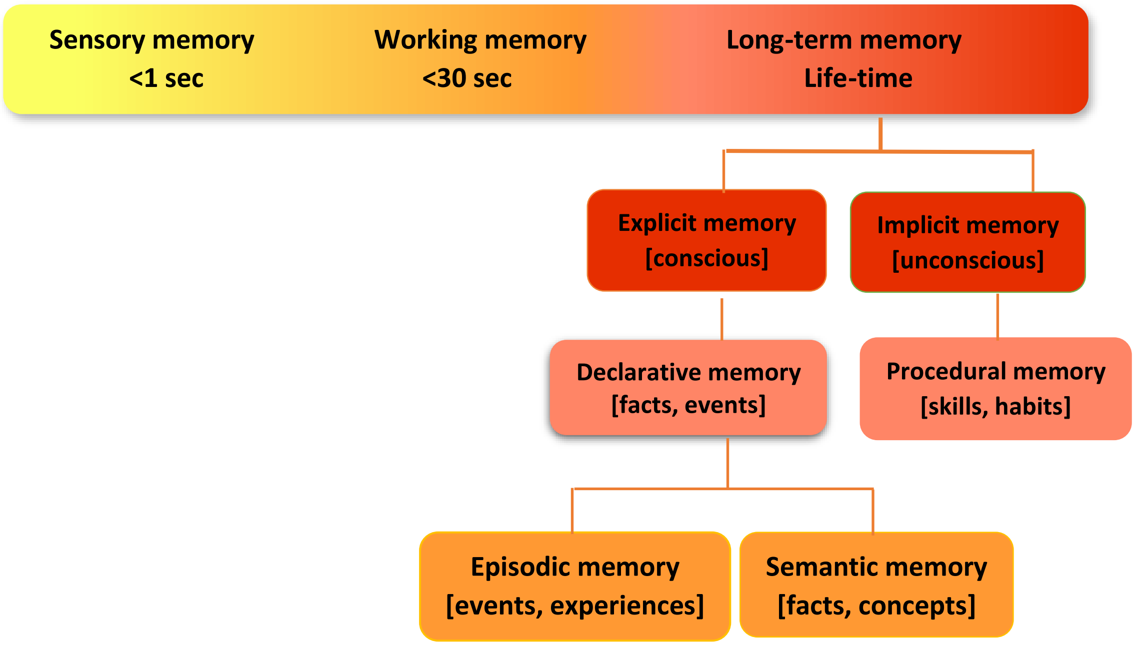


Figure 2. The multi-store model of memory according to memory duration (based on Divjak (2019: 105))

Sensory memory is the shortest in duration; however, it is necessary to retain sensory impressions after the stimulus has passed to allow cognitive processing prior to conscious access (Divjak 2019: 105). The next two types of memory systems are working and long-term memory, which are crucial for the process of learning. The different types of memories listed under long-term memory are classified according to the type of information they store, i.e., declarative and procedural, which are also referred to as explicit and implicit.

According to Divjak (2019), what is now referred to as working memory was earlier called short-term memory. Baddeley et al. (2020) use the term short-term memory to refer to performance on a particular type of task, involving the simple retention of small amounts of information either immediately or after a short delay. They propose that the memory systems responsible for short-term memory form part of the working memory system. According to Baddeley et al. (2020), the difference between short-term memory and working memory is that working memory not only stores information but also manipulates it and allows people to perform complex activities such as reasoning, learning, and comprehension. In essence, short-term memory is part of the working memory system but not equal to the concept of working memory.

Understanding working memory is essential because it is the system where the currently produced or processed information is stored. One of the most well-known models of working memory is found in the work of Baddeley & Hitch (1974: 315). Their model has three components:

(1) visuospatial sketchpad
(2) episodic buffer
(3) the phonological loop

The phonological loop is assumed to have two subcomponents. The first one is responsible for information storage; the second is the articulatory rehearsal processes. The two subcomponents support each other because the phonological loop's storage capacity is limited to a few seconds; however, thanks to articulatory rehearsal, the information can be refreshed and stored for a slightly longer period. Baddeley et al. (2020) present two hypotheses about the necessity of the phonological loop for the human memory system. The first theory is that the phonological loop evolved to aid language comprehension, especially in long sentences, where it is necessary to hold information until the end of the sentence in order to comprehend it. The second hypothesis claims that the system evolved to help us learn the language. This second hypothesis is supported by several experiments where the participants' phonological loop was disrupted, and it greatly interfered with their capacity to learn new words in a foreign language (Papagno et al. 1991). The visuospatial sketchpad is another component of the working memory model, which stores and processes information in a visual or spatial form. The visuospatial sketchpad is likely to play an essential role in keeping track of where we are in relation to other objects as we move through our environment. The sketchpad also displays and manipulates visual and spatial information held in long-term memory. The last component in their working memory model is the episodic buffer. Baddeley & Hitch (1974) defined the episodic buffer as a storage system that can hold about four chunks of information in a multi-dimensional code, allowing various subcomponents of the working memory to interact with long-term memory (Baddeley & Hitch 1974; Baddeley 1992; Baddeley 2007; Baddeley et al. 2020).

In simple words, memory consolidation is the transfer of memories from working memory to long-term memory. That is to say, our working memory capacity greatly influences how fast we can learn and how many items we can learn. Long-term memory is where everything we know, be it words, collocations, valency patterns and constructions, is stored. Due to the versatile types of information stored in long-term memory, it is usually classified according to the kind of information stored there. The main distinction is traditionally drawn between procedural and declarative memory (Llompart & Dąbrowska 2020; Ellis 2008). Declarative memories can be consciously retrieved, while procedural memories consist of different motor and perceptual skills, conditioned behaviours, and automatic processes (Divjak 2019). Ullman (2004) suggests that both systems are vital for language processing. Declarative memory activates knowledge about words and sounds, while procedural memory helps to combine these elements into more complex structures. These two systems build a dynamic system and should be seen as complementary rather than separate systems. What is more relevant to the current dissertation is how these elements get to the different memory systems in the first place. The answer to the question is the above-mentioned process of memory consolidation.

Memory consolidation can take place on two different levels: on the molecular and cellular levels and the system level (Takashima & Bakker 2017). Consolidation at the cellular level occurs in seconds to hours after learning. Consolidation refers to the processes that result in the formation of new synapses or the updating and restructuring of existing synaptic connections. In other words, the processes described by the Hebbian learning model. On the other hand, system consolidation involves brain regions rather than individual neurons and occurs between hours to years after the initial learning process. For this dissertation, memory consolidation on the system level is of particular relevance. Squire et al. (2015: 2) describe in detail the processes occurring during system consolidation:

> Systems consolidation is typically, and accurately, described as the process by which memories, initially dependent on the hippocampus, are reorganized as time passes. By this process, the hippocampus gradually becomes less important for storage and retrieval, and a more permanent memory develops in distributed regions of the neocortex. The idea is not that memory is literally transferred from the hippocampus to the neocortex, for information is encoded in the neocortex as well as in hippocampus at the time of learning. The idea is that gradual changes in the neocortex, beginning at the time of learning, establish stable long-term memory by increasing the complexity, distribution, and connectivity among multiple cortical regions.

In essence, memory consolidation is the process of fixation of information for later use. Entrenchment as a process serves the same goals of encoding and storing information. Takashima & Bakker (2017) have highlighted that the term consolidation originally only referred to the strengthening and stabilisation of memories. The definition of the term evolved, and it started to be used to indicate the processes related to the abstraction of common features, in other words, generalisation. The complementary learning system model[12] assumes that novel memories are not only stored as separate memories but are also integrated into existing knowledge through interleaved learning. The novel memory trace is transformed from a single episodic memory into a generalised semantic representation (Takashima & Bakker 2017). This process helps language learners generalise over language rules[13] and patterns and use them across different communicative situations[14]. When learners are exposed to a new word or word combination, the hippocampus initially encodes an episodic representation of this word. This word can be retrieved from memory; however, it does not interact yet with the existing mental lexicon. Only hours and days later, neocortical connections between the newly-learned word's components, such as phonology, spelling, and meaning, and between the new word and the related existing knowledge is gradually strengthened (Takashima & Bakker 2017). Just as in the case of entrenchment, frequency of occurrence and usage intensity support memory consolidation. The brain mechanisms that underlie entrenchment in memory are influenced to a considerable degree by repeated experience, be it an overt experience or silent rehearsal. This implies that when discussing the mechanisms of learning new and strengthening extant linguistic knowledge, entrenchment is seen as a process that involves the encoding, storage, updating, and retrieval of memories. Electrophysiological data support this idea as it provides evidence that novel words start to behave as existing words after consolidation; thus, they can be recalled and explained what they mean and integrated into larger units like sentences. In an EEG study by Bakker et al. (2015), newly learned words showed pseudo-wordlike N400 responses immediately after training; however, hours after the training, these newly learned words were significantly more word-like, which implies that the process of consolidation has already started.

It is an accepted fact that language is one of humans' most complex cognitive abilities, and we are still far from understanding how linguistic memories are consolidated, and whether grammar and lexicon rely on different consolidation

[12] For the description of the model see Takashima & Bakker (2017) and Divjak (2019)

[13]Rule in this context are used in sense of Langacker (2008: 23). He defines rules as the characterization of some pattern. In cognitive grammar, rules take the form of schemas. Schemas refer to abstract templates obtained by reinforcing the commonality inherent in a set of instances. These schemas are abstracted from occurring expressions and once established as units they can serve as templates guiding the formation of new expressions on the same pattern.

[14] For detailed review of how declarative and procedural memories are formed see Divjak (2019: 99–128)

processes[15]. Takashima & Bakker (2017) suggest that reactivation and rehearsal of linguistic experience not only strengthen individual memory traces but also enable the consolidation of neocortical connections between the representations. The integrating process underlies the formation of collocations, prefabs, and generalised abstract patterns and constructions.

In conclusion, it can be said that memory consolidation is one of the key processes of entrenchment. Every entrenched linguistic item and structure must be stored in our long-term memory. The degree of entrenchment represents the strength of those stored memories. Memory encoding and consolidation are dynamic and gradient processes influenced by the type of input and the cognitive capacities of the language users.

4.2.2 Chunks and chunking

The importance of chunking for the process of entrenchment is anchored in the original definition of the concept by Langacker (1987:59): "With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit". In this definition, chunking is defined as a unit[16] formation process, i.e., a process during which individual pieces of information are grouped together in a meaningful whole.

Chunking, like entrenchment, is a multifaceted concept. Chunking can be deliberate and automatic. Deliberate chunking is most often regarded as a way to bypass the limits of short-term memory, and automatic chunking is more related to the processes that are relevant for encoding the linguistic structures in long-term memory (Divjak 2019: 142). Deliberate chunking is present both in oral and written comprehension. Chunking the input facilitates the listener and the reader to process larger stretches of speech. Clause and phrase boundaries are potential candidates for chunk

[15] According to the usage-based approach to language and construction grammar, there is no sharp division between grammar and lexicon. Some of the linguistic items are more lexical and others are more schematic. The main question is whether lexical linguistic items are consolidated together with the schematic patterns in which they are frequently used or these patterns emerge later from usage and input. Thus, with grammar and lexicon the distinction between more lexical and more schematic linguistic items is meant.

[16] Unit is defined by Langacker (2008: 216) as "recurring aspects of processing activity. To different degrees, these patterns of neural processing have coalesced as entrenched cognitive routines that can be activated whenever needed. Linguistic units are dynamic in nature, residing in aspects of cognitive processing." It is through occurrence in usage events that linguistic units arise in the first place. More precisely, they are abstracted from usage events through reinforcement of recurring commonalities Langacker (2008: 458). Thus, language can be describes as a "a structured inventory of conventional linguistic units. This structure - the organization of units into networks and assemblies is intimately related to language use, both shaping it and being shaped by it." Langacker (2008: 222) . Language is viewed as a collection of conventional linguistic units. They are conventional because they represent established linguistic practise in a certain speech community.

boundaries, just as punctuation marks in written texts and pauses/intonation in oral input (Sinclair & Mauranen 2006). On the lexical level, idioms, collocations, and different phraseological units tend to stick together and thus are usually processed as one structural unit and could also possibly indicate chunk boundaries in longer stretches of speech. Automatic chunking leads to the reorganisation of mental representations into holistic units that are later retrieved from memory in one single step. Automatic chunking, thus, is more relevant for entrenchment than deliberate chunking[17].

The father of the chunking theory is Miller (1956). He was the first researcher to propose an information measure for cognitive systems based on the concept of a chunk. According to Miller (1956), each chunk collects a number of pieces of information from the environment into single units. In other words, information is being compressed to require less storage space in the cognitive system and overcomes this way its storage space limitations. Another frequently cited definition of chunks comes from Newell (1990: 7):

> A chunk is a unit of memory organisation, formed by bringing together a set of already formed chunks in memory and welding them together into a larger unit. Chunking implies the ability to build up such structures recursively, thus leading to a hierarchical organisation of memory. Chunking appears to be a ubiquitous feature of human memory.

It is undeniable that chunking and memory are inseparable. Both memory and chunk formation are triggered by repetition. Chunking is a property of both production and perception and contributes significantly to fluency and ease of processing (Bybee 2010). The higher the degree of entrenchment of a chunk, the more fluent its execution is, and the more easily it is comprehended. In production, the effect of chunking can be observed in the overlap and reduction of articulatory gestures. In perception, the effect is visible in anticipation of what is coming next. Within this dissertation, these two effects are treated as signs of automaticity, which is the end product/phase of entrenchment and will be discussed in the next section. Moreover, chunking is the process behind the formation and use of formulaic and prefabricated sequences of words, and additionally, it is the primary mechanism leading to the emergence[18] of

[17] Gobet et al. (2001) refers to these two different types of chunking as goal-oriented and perceptual chunking. The former is a deliberate and conscious control of chunking processes, and the latter is an automatic and continuous process of chunking during perception. An example of perceptual chunking is how letters are grouped into words, words into phrases, and then into sentences. This leads to changes in the memory system and language production, and perception. Latencies for chunks are shorter than for individual elements.

[18] According to Langacker (2008: 220), units emerge via progressive entrenchment of configurations that recur in a sufficient number of events to be established as cognitive routine. Once these routines emerge, they function as templates in constructing and interpreting new expressions.

constructions (Bybee 2010). What is interesting to highlight is that the units that are formed during automatic chunking are the units that we use as clues during deliberate chunking.

Besides playing such a vital role in language comprehension and perception, chunking plays a key role with respect to learning (GOBET et al. 2001). During the process of chunking, associations on the syntagmatic level become stronger, while on the paradigmatic level, they become weaker (Bybee 1985, 2007b; GOBET et al. 2001). The strengthening of the syntagmatic associations can be interpreted from two angles. First of all, when we refer back to the Hebbian learning model, association strength can be viewed as the formation of neuronal assemblies, where the synaptic connections between single neurons become stronger, and activation of one of the neurons triggers activations in the related neurons. This angle represents the neurobiological bases of chunking. The second perspective is represented by the dynamic framework of Schmid (2020) and Ellis (2017). Ellies (2017:117) defines chunking as the development of permanent sets of associative connections in long-term storage. According to Schmid (2020), chunking leads to the strengthening of syntagmatic associations. Many of these definitions highlight that chunks are defined as certain types of associations. These associations can be interpreted to be between form and meaning or between words and specific patterns[19]. The syntagmatic associations are defined by Schmid (2020) as collocations and other co-occurrences[20] in the language. His view of syntagmatic connections is identical to Bybee's (2010) view on the effects chunking has on language use. Thus from a psychological point of view, chunking is a perceptual and conceptual information grouping process fostered by the language users' capability to track transitional probabilities between frequently co-occurring elements in both oral and written language (Schmid 2017c).

According to Christiansen & Chater (2016), chunking is the process that allows humans to deal with the fleeting character of language and memory. We continuously receive input during conversation or reading, and our brain has to deal with this amount of information and make it available for perception or production almost immediately. To deal with this "Now-or-Never" bottleneck, the brain must compress and recode linguistic input as rapidly as possible (Christiansen & Chater 2016). Besides compressing linguistic input, the language system must deploy all available information predictively

[19] For an extensive review of these associations on different linguistic levels consult Ellis (2017)

[20] For instance, when particular verbs trigger certain valency construction patterns, e.g. *give, offer, throw* are often observed in the ditransitive construction.

to ensure that processing is efficient and smooth. According to this view, chunks not only make language perception and production more efficient, but they are necessary to perform online processing. The strong relationship between chunking and language processing was empirically confirmed by the results of a study by McCauley & Christiansen (2015). The results have shown that the chunking skills of participants predicted their online sentence processing abilities. This result is consistent with the notion that chunking at higher levels may reduce the computational demands involved in processing embedded clauses.

The result of the automatic chunking process is the creation of holistic units[21], which are accessed and retrieved as one piece of information, thus barring, or at least impeding, access to its component parts. These holistic representations are formed by repeated chunking in working memory and subsequent entrenchment of the elements in long-term memory. As pointed out by Blumenthal-Dramé (2012: 76), holistic chunks possess the following features:

a) *pattern completion*: the component parts of a given chunk will evoke the whole;
b) *emancipation*: the perception of the whole will be autonomous from that of its parts;
c) *top-down coercion*: the parts will be perceived and interpreted in the light of stored knowledge of the whole (rather than vice versa);
d) *ease of memory*: it will be easier to remember chunks than non-chunked bits of information.

In conclusion, it can be said that chunks and memory are strongly related. Chunking information makes language production and processing cognitively less demanding. Repetition[22] is the strongest trigger of chunking, and high frequency in exposure is less critical for the process of chunk formation than the actual rehearsal and repetition of the chunks. Furthermore, chunking is subject to the Power Law of Practice, which stipulates that performance improves with practice, but the amount of improvement

[21] Siyanova-Chanturia (2015) has a critical view on the topic of holistic processing. She argues that most of the studies in psycholinguistics neglect studies carried out in the field of psychology. Her main argument contends that most studies claim that their holistic storage and processing is a "psychologically valid claim"; nevertheless, psycholinguistics research is not drawn on the existing body of evidence coming from recent empirical studies from the field of psychology. Siyanova-Chanturia's (2015) second claim is that sometimes linguists try to rely solely on corpus data to gather evidence for holistic storage. She claims that the studies that associated formulaic expressions, which are processed more accurately and faster, with holistic storage can only claim that these expressions have a processing advantage caused by their frequency, which leads to higher familiarity and predictability. In essence, a processing advantage observed for formulaic sequences over novel phrases in reading and production studies is a vital piece of evidence that argues for the crucial role of phrasal frequency in language acquisition. Still, it does not automatically mean that these phrases are stored holistically.

[22] Repetition and frequency are not equal in this context. Repetition refers to the usage intensity and not only the mere corpus-based frequency.

decreases as a function of increasing practice of frequency. Thus, when the chunking status is achieved after several repetitions, further repetition will maintain the chunk status but will not further increase it. This is also true of entrenchment; when the highest degree of entrenchment, namely automaticity, is achieved, usage intensity and frequency will only help maintain this status but will not further increase it.

4.2.3 Routinisation and schematisation

These two final processes associated with the process of entrenchment refer to the anchoring of the learnt items into long-term memory, their abstraction and generalisation. Both of these processes are the most exhaustively discussed in the work of Schmid (2015, 2017a; 2020). In his earlier work (Schmid 2015: 13), he treated routinisation and schema formation as two related processes and defined them as cognitive and neural effects of the activation of repeated identical or at least similar patterns of associations. Routinisation is the process responsible for strengthening and increasing the automatisation of associations. On the other hand, schematisation derives second-order associations from the commonalities of first-order associations. Moreover, schematisation is regarded as a byproduct of statistical learning and routinisation (Frost et al. 2015: 118). This definition of schematisation stems from Langacker (2008: 17). For Langacker, schematisation was a way to arrive at higher levels of abstraction. Schematisation, associations, automatisation, and categorisation are the four independently existing cognitive processes that are the driving forces of language acquisition, development, and usage. Schematisation comes into play as soon as frequently occurring routines contain some variable slots that can be filled with different items. Schematisation enables the learners to generalise over the items that have been observed in different slots[23] and to arrive at abstract schemas. Moreover, schematisation can be regarded as a key component of entrenchment because different abstract schemas lie at the heart of language learning and speakers' capacity to form new sentences that have not been experienced in the input (Schmid 2017c). While chunking allows language learners to acquire collocations, idioms, and other lexicalised multiword units, schematisation is inevitable for the different construction and valency patterns to emerge. Token

[23] Slot is defined a position within a valency pattern / construction that can / must be filled by a particular linguistic item. E.g. the ditransitive construction has three slots that correspond to the *giver*, the *givee* and the *thing given*. For a detailed description see Herbst & Schüller (2008) & Goldberg (1995); Goldberg (2003, 2006, 2019).

frequency and repetition, together with statistical learning, lead to the formation of chunks. On the other hand, type frequency is the main driving force of schematisation. It promotes abstraction and allows schematisation to take place (Bybee 1995).

It is often argued that routinisation is the process that leads to automatisation (according to the definition of Langacker (2008: 16–17)). However, Schmid (2017) argues that automatisation is not an all-or-nothing process; some mental activities are more automatic than others, and their automaticity depends on a large number of factors besides repetition and frequency. Therefore, Schmid (2017) proposes to talk about more or less highly routinised linguistic elements and constructions. These two processes can be treated as complementary and not only as exclusive. Within the framework of this dissertation, routinisation is defined as a process that, along with the other processes discussed in this chapter, leads to entrenchment. Entrenchment as an end-product is defined by the achievement of the stage of automaticity; thus, when the processing, accessing, and production of linguistic elements, patterns, and constructions are performed rapidly, without much cognitive effort or control. This automaticity is, of course, gradable. Lexical items are likely to exhibit automatic processing and production, but constructions and other abstract patterns are just as likely to be automatised. Entrenchment, in general, can also be treated as a gradient state, which is even part of the original definition (Langacker 1987: 59) and is referred to as the degree of entrenchment.

4.3 Conclusions on entrenchment as a process

Emerging from the discussion above, it can be said that entrenchment is a complex, multifaceted process. The process includes the generation and updating of linguistic representations, their encoding in memory, chunking and schematisation. The generation of the representations is driven by the different learning strategies that language users use to acquire language in general and specific linguistic items in particular. Learning manifests itself in the reorganisation of the neural networks that allow us to produce and comprehend speech so effortlessly. The newly learned and strengthened representations are, in the next step, encoded in memory for later usage. To get the most out of our memory, these representations are reorganised into chunks, which not only increases our working memory capacity but also makes language comprehension and production smoother and more effortless. The final stage of the process is schematisation, which allows language users to entrench not only lexicalised items but also abstract units like constructions. Figure 3 below attempts to illustrate the potential order of the processes that convert linguistic input into stored mental representations.

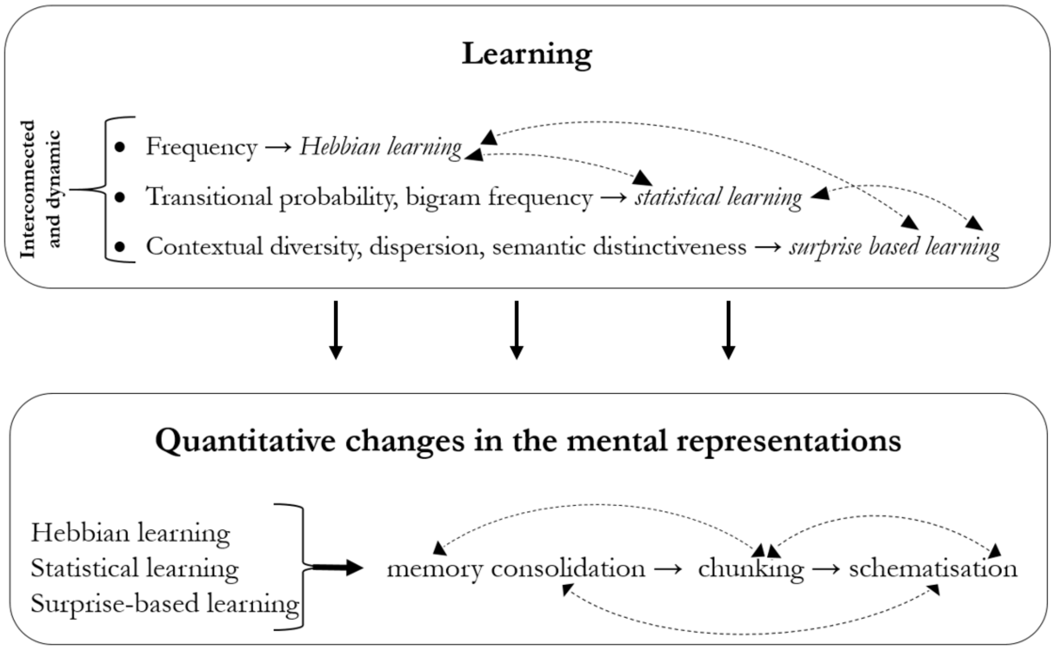


Figure 3. Processes that contribute to entrenchment

The upper part of the figure demonstrates that the three learning processes are not separate processes, but they are dependent on each other and form a dynamic network. The order in which the learning processes are listed implies the order in which they are likely to happen during entrenchment. Moreover, the possible linguistic metrics that trigger these processes are indicated.

In the initial stages of language acquisition, language users store single words[24]. The choice of which words to store largely depends on the frequency of these words in the input. As a child's mental lexicon gets richer, they begin noticing the underlying patterns that exist in the input. This noticing is achieved through statistical learning. As pointed out above, statistical learning is a psychological phenomenon, while Hebbian learning attempts to connect this psychological phenomenon with the neurological

[24] By word a mono- or multimorphemic linguistic item is meant that exist and an informational unit in language. According to Tomasello (2005: 72–73), children's early linguistic communication consists not of words but of utterances. By utterance he means the smallest unit with which a person expresses a complete communicative intention. The form of these utterances can be simply a single adult word expressed with a certain emotion. That is to say these words are already, in a very simple way, composite structures. The question is whether they store single words and use them in different communicative situations with different intentions, or whether each word/utterance is stored together the most often used communicative intention. According to Tomasello (2005: 75), children start language acquisition and production with whole, meaningful utterances. Later children extract words (with their functions) from utterances. It is undoubtedly the case for function words.

underpinning of the learning process. Statistical learning enables language users to notice the patterns, and Hebbian learning enables them to register these patterns and encode them into neural assemblies. Surprise-based learning can only occur if a sufficiently large number of such patterns are encoded into our memory because the surprisal effect emerges only if our expectations about the input are not met. In other words, language users rely on predictive processes when processing the input, and if an unexpected word or pattern is observed in the input, they will potentially register this pattern. It has to be highlighted that these learning processes are interconnected, and they influence each other. The dashed lines in the graph indicate that they impact each other and that some of these processes might be more prominent for certain linguistic structures. Thus, both statistical and surprise-based learning strengthen the connection between the neurons, i.e., they facilitate Hebbian learning, and Hebbian and statistical learning consequently strengthen the effect of surprise-based learning.

The second part of the figure aims to show the outcome of these learning processes – the quantitative changes in the mental representations. First, these newly learnt linguistic items are encoded into memory (memory consolidation), then memories that are frequently activated together begin to form holistic units (chunking), and finally, abstract rules and patterns emerge from these stored representations (schematisation). Within the framework of this dissertation, it is suggested that chunking cannot take place without the members of the chunk being stored in memory. Chunking is seen as the reorganisation of mental representations, i.e., connecting linguistic items that are frequently used and observed in the input into one neural assembly. Chunking during perception occurs because language users rely on previously stored memories. Thus, when encountering a longer stretch of speech, they will chunk it, relying on the stored phrase boundaries, collocations, or punctuation. Without this prior knowledge, they would not be able to organise the input into meaningful groups. As indicated by the dashed line, these chunks respectively influence stored memories and the process of memory formation. The final process indicated in figure 3 is schematisation; it is the process of moving from more lexical items to more schematic ones. In other words, schematisation helps language users to achieve abstractions[25]. These schemas can be "recruited and exploited to produce

[25] The claim that schematisation follows chunking is indirectly supported by a number of research papers. Among them,

a) Ellis (1996: 92) who suggests that language learning is the learning and analysis of sequences. The learner must acquire sound sequences in words. The learner must acquire word sequences in phrases.

and comprehend new utterances" (Schmid 2018: 168). Schematisation can possibly be regarded as a necessary process to form constructions and grammatical rules in the language. Once again, it has to be emphasised that the order of these processes is only hypothetical, as there is currently no method that would allow us to validate the sequence. The processes were arranged in this order following the logic described above. All these processes jointly lead to automaticity, which is seen as the final stage of entrenchment in this dissertation. A quote from De Smet & Cuyckens (2007: 188) very neatly summarises the processes that are associated with entrenchment: "a highly entrenched unit represents an automated, routinized chunk of language that is stored and activated by the language user as a whole, rather than 'creatively' assembled on the spot". The definition emphasises that entrenched items have to be stored in memory, they are most likely to be organised in the form of chunks, and these chunks are automatically recalled and processed when observed in the input or when selected for production.

These sequences form the database for the abstraction of grammar. In other words, with the help of statistical learning we learn the sequences that frequently occur in the input and later these sequences (chunks) serves as the bases for generalisation and schematization. In a later article, he is even more specific about it: "A chunk is a unit of memory organization, formed by bringing together a set of already formed chunks in memory and welding them together into a larger unit. Chunking implies the ability to build up such structures recursively, thus leading to a hierarchical organization of memory" Ellis (2003: 87)

b) Blumenthal-Dramé (2012: 1) writes that highly entrenched multi-word chunks are the stepping stones to the formation of abstractions at different levels of analysis. The above used quotation also supports the proposed order of the processes that take place during entrenchment.

5 Entrenchment as a product

According to the framework of this dissertation, the final stage of entrenchment is signalled by automaticity. This is the stage when the neural assembly for the representation of a linguistic item is generated, it is rehearsed to the extent that the threshold level for its activation is low, and its processing and production are done fast and effortlessly without demanding much attentional capacity or consciousness. Automaticity is present both in production and comprehension and can take different forms. It is envisaged as the last stage of entrenchment because only stored memories can be automatised since the linguistic items first have to be extracted from the input and learned. That is to say, automaticity emerges from language use and production for which input and different learning mechanisms, together with other cognitive processes (e.g., memory consolidation, chunking, schematisation), are necessary.

5.1 Automaticity

Most people have a clear idea of what automaticity means in a very general sense. Some doors slide open automatically, as in a supermarket or a lift, and most people have either heard of or even driven a car with an automatic gear shift. Moreover, most people have some automated skills such as cycling, typing blindly, or driving to work along the same route every day. These skills are automatic because they function correctly without requiring any cognitive or physical effort on our part. One of the most automatised skills we have is producing and comprehending language. Through a complex chain of mental processes, our thoughts and feelings are converted into oral or written speech and incoming sound waves and sequences of letters are transformed into meaningful units. Effortlessly and with stunning speed, we access and retrieve linguistic information from our long-term memory and "apply lower-level and higher-level cognitive abilities such as perception, attention, categorisation and inferencing while producing and comprehending utterances" (Schmid 2014: 2).

This section aims to describe how automaticity is conceptualised within the framework of this dissertation and highlight how this conceptualisation differs from or aligns with the existing definitions.

The guiding principles to conceptualise automaticity within this dissertation are the following:

a) Automaticity is envisaged as a manifestation of entrenchment;
b) Linguistic input triggers different learning processes (Hebbian, statistical and surprise-based learning) that lead to the encoding of new information in long-

term memory and that strengthen stored representations of already existing memories;

c) Stored memories are reorganised into chunks or are schematised depending on their structure, i.e., whether they are on the more lexical or schematic end of the lexico-grammatical space[26].
d) With repetition, representations of linguistic structures are strengthened to the extent that their activation and production become automatic. In other words, the activation threshold sinks, and it does not require much cognitive effort to comprehend the linguistic structure or select it for production;
e) This automaticity is manifested in fluent comprehension and production;
f) Fluency in comprehension is signalled by efficiency, lack of deliberate control and attention;
g) Fluency in production is signalled by the development of neuromotor routines that lead to the effortless execution of articulatory gestures. The effortless execution of these gestures results in temporal and phonetic reduction;
h) Automaticity is conceptualised as a gradual process which implies that the absence or presence of an automaticity feature (efficiency, attention, awareness, control) cannot be used to draw conclusions about the whole process;
i) Automaticity features can compensate for each other, and the joint effect of these features determines the degree of automaticity;
j) Besides efficiency, awareness, attention and control, stimulus intensity, expectations, novelty, goal relevance, recency, salience, familiarity and frequency influence whether the process is executed in a more or less automatic manner;
k) Moreover, the empirical evidence suggests that the degree of automaticity of a process is influenced by the language user's cognitive skills and language experience. Among the cognitive skills, processing speed and statistical learning skills are expected to influence the ease with which automaticity is achieved.

[26] The lexical end of the space is represented by collocations, fixed expressions and other multi-word expression. The grammar end of this lexico-grammatical space is exemplified by constructions and valency patterns that are abstract patterns e.g., the ditransitive or caused-motion construction.

5.1.1 Automaticity as a process of strengthening mental representations and associations

In cognitive psychology, automaticity has a history of more than half a century. One of the first theories on automaticity was formulated by Schneider & Shiffrin (1977a), who proposed a two-process theory of human information processing. These two processes are defined as *automatic* and *controlled*.

> "Automatic processing is activation of a learned sequence of elements in long-term memory that is initiated by appropriate inputs and then proceeds automatically—without subject control, without stressing the capacity limitations of the system, and without necessarily demanding attention" (Schneider & Shiffrin 1977a: 1).

Automatic processes are contrasted with controlled processes, which are defined as

> "temporary activation of a sequence of elements that can be set up quickly and easily but requires attention, is capacity-limited (usually serial in nature), and is controlled by the subject" (Schneider & Shiffrin 1977a: 1).

The claim that automatic processes represent learned sequences and that an ample amount of practice is necessary to achieve this state is in agreement with the conceptualisation introduced in this dissertation. Language users have to be exposed to the input; they have to engage with it actively (e.g., process or produce the linguistic structures) in order to learn and encode it into memory. A further similarity between these two conceptualisations of automaticity is the reorganisation of the stored memories into larger units: chunks and schemas according to the framework of the current dissertation and nodes according to the theory of Schneider & Shiffrin (1977a). According to Schneider & Shiffrin (1977a), stored memories are complex nodes that build a network and interrelations among nodes. Each node can contain a complex set of information units and their associations with other nodes. This description very much resembles the neural assemblies that are the bases of Hebbian learning, and this learning is regarded as one of the processes that lead to automaticity within the framework of this dissertation. In other words, neurons that are frequently activated together form a network. The automatic activation of these nodes is the result of their interconnections, i.e., once a member of the node is activated, it triggers a chain reaction, and all the elements within that node become automatically activated. The ease with which these nodes are activated largely depends on their activation threshold level. The formation of chunks is envisaged in this dissertation in a very similar way. The nodes are stored representations that, with sufficient practice, are reorganised into larger units, i.e., chunks. The more often the items within a chunk are activated together, the lower their activation threshold will be. This threshold level refers to the amount and the strength of stimuli needed to trigger the retrieval of these linguistic structures from long-term memory. Hebbian learning also postulates that stronger neural assemblies have a lower level of activation. The level of threshold decreases with practice, i.e., the more often the linguistic structure is activated, the lower the threshold is. Practice can be quantified as the frequency of occurrence of the structure in the input and its frequency of use in production.

Besides these similarities, there are a number of points where the two conceptualisations of automaticity diverge. Schneider & Shiffrin's (1977a) theory proposes that automatic processes cannot be modified, i.e. automatic processes are not subject to any changes. This implies that automatic processes cannot become less automatic due to disuse. This claim goes against the core idea of entrenchment and automaticity outlined in this dissertation: entrenched memories are gradual, and the strength of the entrenched memories depends on their frequency of use. Therefore, extended periods of disuse negatively influence the degree of entrenchment because these entrenched memories are constantly updated and extended, and their representation's strength directly correlates with their usage intensity. This conceptualisation of entrenchment in general and automaticity, in particular, is supported by the idea of Hebbian learning, which also highlights the need for constant repetition: "neurons out of sync delink" (Schmid 2020: 207).

Moreover, within the theoretical framework of this dissertation, it is proposed that automatic processes can be transferred to novel situations. Schneider & Shiffrin (1977b: 156) claim that controlled processes (in comparison with automatic ones) require short-term memory capacity, but this limitation is balanced with the fact that controlled processes may be set up easily, altered and applied to novel situations for which no automatic sequence has been learnt. This claim suggests that any automatised knowledge stored in procedural memory is completely useless when a novel situation is observed. According to the framework outlined in this dissertation, schematisation is the process that allows language users to form abstract patterns that are extended to new situations. In other words, if someone wishes to express the idea of caused-motion[27] and they have entrenched the caused-motion construction but have witnessed it mainly with the verb *put*, they would still be able to find other verbs (such as *break*, *throw*, *push* etc.) that can express the same idea. Furthermore, the person would be able to comprehend a sentence that uses so far unwitnessed verbs

[27] The basic sense of caused-motion is that a causer or agent directly causes a thing to move a new location.

in this construction without much cognitive effort, e.g., *Pat sneezed the foam off the cappuccino*.

The current dissertation has the objective, besides conceptualising entrenchment and automaticity, of empirically validating this conceptualisation. For this reason, it has to be highlighted that Schneider & Shiffrin (1977a, 1977b) formulated their theory on empirical evidence. However, their empirical evidence stems exclusively from a serial search task[28] and the performance of 4 participants. A theory that hopes to arrive at reasonable conclusions about such a complex phenomenon as automaticity should consider using a wide range of tasks that manifest automatic processing and test participants with different social and educational backgrounds and cognitive skills.

The most controversial issues related to Schneider & Shiffrin's (1977a) theory of automaticity concern the character of automatic processes (principle h from the list above): whether automaticity is an all-or-none phenomenon or gradual. This is directly related to their claim that automatic processes are hard to modify, alter or suppress. In other words, the main question is whether automatic processes are in a constant state of flux or remain stable once the state of automaticity is achieved. As was highlighted above, this dissertation envisages automaticity as a gradual process. This assumption emerges from the conceptualisation of entrenchment outlined in the previous two chapters. Since automaticity is seen as the end stage of entrenchment, this line of reasoning can be used to conceptualise automaticity also as a gradual process.

The majority of theories that conceptualise automaticity likewise disagree with the neat division between automatic and controlled processes and advocate for the idea of treating automatic processes as having a gradual nature. Needless to say, these theories do not conceptualise automaticity alike, but their core principle is that automaticity is not an all-or-nothing phenomenon.

Cohen et al. (1990) and Anderson (1992) emphasise the importance of strengthening neural pathways and the procedures[29] responsible for executing any cognitive processes. Every process that is initiated in our cognitive system can be viewed as a sequence of connected modules. These connected modules form certain neural pathways. The degree of automaticity is determined by the pathways and the strength of these pathways. The execution of every cognitive process requires the

[28] A serial search task is a visual search task that focuses on identifying a previously requested target surrounded by distractors possessing no distinct features from the target itself.

[29] *Neural pathway* is a term used by Cohen et al. (1990) and *procedure* or *algorithms* is used by Anderson (1992).

presence of such a pathway that allows activation in relevant sensory modules. The speed and accuracy of the execution of the process largely depend on the speed and accuracy with which information flows from one of these modules to the other. If a process is repeated frequently enough, the sequence of steps needed to perform the process is stored in procedural memory, and as a result, the stored memory sequence can be retrieved automatically. The main driving force of a process to become automatic is practice.

Both of these theories (Anderson (1992) and Cohen et al. (1990)) rely on the strengthening of stored memories. In this respect, these two theories share some similarities with that of Schneider & Shiffrin (1977b, 1977a). The principles used to conceptualise automaticity in this dissertation likewise emphasise that strengthening the representations of items is crucial for a process to become automatic. Thus, if a process becomes automatic, something has to be strengthened, either the connection between the nodes, the pathway, the algorithms, or the mental representations. Moors (2016) similarly defines strengthening as the transition of processes from less automatic to more automatic. This view agrees with the conceptualisation of automaticity outlined in this dissertation because this transition involves the shift from algorithm computation (multi-step memory retrieval) to single-step memory retrieval. This shift can be described in the following way:

> After sufficient repetition of the same chain of steps going from the same input to the same output, a direct association is formed between the input and the output, such that the presentation of the input alone directly activates the output (Moors 2016: 279).

The definition provided above emphasises the crucial role repetition plays in this transition. In other words, automaticity cannot develop without repetition, just as there is no entrenchment without repetition. This is a further point that unites the theories outlined so far despite their different approaches to the conceptualisation of automaticity.

A further theory that conceptualises automaticity as the process of strengthening mental representations is developed by Schmid (2020). However, he refers to the process that leads to the smooth execution of different processes resulting from extensive practice as *routinisation*. The term *automaticity*, according to him, does not provide a good enough definition of the processes that define entrenchment; that is why he claims that we language users rely on more or less routinised processes. In other words, what becomes entrenched and in what way largely depends on the repeated processing of patterns of associations and on what is already entrenched in the language user's associative network (Schmid 2020: 226). Routinisation has already been described in chapter 4.2.3. The theoretical framework of this dissertation sees routinisation as a necessary process for schematisation and automatisation to happen, but routinisation is not superior to automaticity.

The main idea behind routinisation is that it strengthens the associations between linguistic units on different levels. According to Schmid (2020), there are four types of associations: syntagmatic, symbolic, pragmatic, and paradigmatic. Syntagmatic associations are based on the sequential processing of the input and can unfold on different levels. These different levels include: lexical, phraseological, lexico-grammatical and grammatical. The routinisation of symbolic associations affects the link between the mental states representing and controlling the perception or articulation of linguistic forms and the mental states representing their meaning.

Simply put, symbolic associations strengthen the processing of pairings of form and meaning. Pragmatic association, just like symbolic association, targets meaning, but this meaning is context-dependent in that it bridges the gap between entrenchment and conventionalisation by linking cognitive activities with interpersonal activities taking place in certain social contexts. Finally, according to Schmid (2020: 286), paradigmatic associations are "dynamic cognitive substrate of the structural principle of opposition".

These associations help language users identify similarities and differences between usage events and recognise different utterance types. These four different types of association, along which routinisation can happen, agree with the principles that outline automatisation in this dissertation. The only difference between them is that Schmid (2020) explicitly defines that routinisation can happen on these four dimensions, and the theoretical framework of this dissertation establishes that this strengthening can happen at any point in the lexico-grammatical space. On the more lexical scale, strengthening leads to the sequential processing of collocations and other multi-word units, while on the grammatical end of the scale, strengthening leads to the emergence of constructions and valency patterns.

5.1.2 Features of automaticity and the role they play in defining automaticity

The second group of theories of automaticity target the features that are associated with automatic processes and question whether these features are crucial for the process to be defined as automatic. Automatic processes in the theory of Schneider & Shiffrin (1977a) are those processes that are triggered by the appropriate stimulus, without the need for the subject's active engagement in the process (*unintentional*), without demanding attentional capacity (*efficient*) or much time (*fast*), and without consciousness (*unconscious*), making it difficult to counteract the process (*uncontrollable*).

Two influential theories on automaticity by Bargh (1992) and Moors (2016) focus on how these features are related to automaticity and define it. These theories postulate that features like attention, control, intention and awareness can signal automaticity, but at the same time, they highlight that the presence or absence of a feature does not define whether a process is automatic or not, as they do not believe in the all-or-nothing nature of automaticity proposed by Schneider & Shiffrin (1977a) which implies that there is coherence among these features. Thus, a process that is unintentional is always efficient, fast, and hard to control in the sense that it is hard to modify or deliberately stop. These four features (attention, awareness, control and intention), according to Bargh (1992) and Moors (2016), do not co-occur perfectly, meaning that not every automatic process occurs without the subject's attention, control, awareness and intention, and because of this, processes manifest different degrees of automaticity. Furthermore, Moors (2016) argues that these features are a subset of many factors affecting language processing. Finally, these factors can compensate for each other; thus, they have a joint effect on the degree of automaticity and the quality of the input's representation.

The present dissertation agrees with the view that features of automaticity should not be defined as an all-or-nothing phenomenon as this would mean that any processing activity would be defined as automatic or non-automatic based on examining if one of these features is present or not (e.g., whether the language user's attention is required or not). Automaticity is a more complex concept, and on certain occasions, an automatic process might require our attention. For instance, when listening to auditory input with poor quality, we deliberately pay attention and seek meaning in the input, and on other occasions, listening happens without our control, e.g., when sitting on the metro and after the ride, we somehow remember some of the ads or have an earworm from a song that was on the radio, though we were not paying attention to these things.

In most of these theories, automaticity is defined as a domain-general feature of information processing. There is no denying that many different processes, like playing the piano or dancing, can become automatic. Langacker (2008) explicitly connects automaticity to language production and processing within the framework of entrenchment. He refers to automaticity in the following way: "through repetition or rehearsal, a complex structure is thoroughly mastered, to the point that using it is virtually automatic and requires little conscious monitoring" (Langacker 2008: 16). It emerges from his description that the main feature of any automatic process is little to no conscious attention or awareness. Once again, this underscores that despite the different conceptualisations of the process of automaticity by different researchers, it is unified in the sense that it relies on the same features, and among these features, attention and awareness occupy a crucial position. In the same manner, as no theory of entrenchment exists that would *not* consider frequency a triggering force behind entrenchment, most theories that conceptualise automaticity in one form or another refer to attention.

Given all that has been mentioned so far, one may argue that despite the criticism that different researchers have expressed about Schneider & Shiffrin's (1977a, 1977b) theory, all theories have some core concepts shared among them. They can be summarised by the fact that a process is automatic if the single steps that are the building blocks of the processes are connected, and the transition from one step to the next does not require a decision to be taken; it can happen without deliberate control, awareness, and attention. The main deviation between the theories stems from these features' necessity, importance, and whether automatic processes are stable or gradual. Table 2 visualises how automaticity has been conceptualised in these different theories. Moreover, it summarises the core principle of each theory.

Table 2. Summary of the theories on automaticity

Year/Author	Novel task	Highly practised task
1977 **Schneider and Shiffrin**	controlled → controlled (limited-capacity, requires attention and flexible)	→ automatic (requires considerable training and difficult to modify)
1990 **Cohen, Dunbar and McClelland**	indirect → (indirect with very weak direct) → weak direct (slow, might consist of a series of steps that could be interfered with, depend on declarative memory, requires attention an allocation of attention)	→ strong direct (fast, less susceptible to intereference, less influenced by allocation of attention, has a strong pathway)
1992 **Anderson**	**Consistency promotes automaticity** -→	highly practised skills with strong production rule (fast, low error rate, decays with disuse, interferes less with concurrent tasks)
1992 **Bargh**	**Conditional automaticity** -→ *If* (x,y,z) → *THEN* (autonomous process)	preconscious automatic postconscious automatic goal-oriented automatic

2016
Moors

Automaticity features are not fixed properties of processes but are rather factors that jointly influence whether process will occur

less automatic → more automatic

the degree of automaticity depends on:
a) factors related to current stimulus (duration, intensity, [un]expectedness, goal, novelty)
b) factors related to prior stimuli (frequency, recency)
c) factors related to prior stimulus representation (goals and expectations)

2020
Schimd

Language consists of more or less higly routinised linguistic elements or constructions

less routinised → more routinised on the
- syntagmatic level
- symbolic level
- pragmatic level
- paradigmatic level

The current dissertation

Frequent repetition triggers learning through which linguistic element become encoded into memory and can be accessed and produced more or less automatically depending on the context, expectations and goals of the language user

driving forces – learning – memory consolidation – chunking and schematisation → more or less automatic process

5.2 Automaticity in language comprehension

Automaticity, just like entrenchment, is a theoretical concept that can only be empirically validated if one relies on some measurable features. When measuring automaticity either in language comprehension or production, one has to predefine a set of features that would signal automaticity. Even if the previous section just concluded that these features do not necessarily indicate the degree of automaticity, one has to rely on certain experimentally measurable manifestations of the process. In language comprehension, the anchor points that are used to validate automatic comprehension include efficiency, the lack of purposeful control, attention, and awareness.

Is automatic comprehension efficient?

When talking about efficiency in language comprehension, the amount of information simultaneously decoded determines the process's efficiency. This amount is usually determined by the language user's working and short-term memory capacities. Frequently co-occurring multi-word expressions (a.k.a chunks, idioms, collocations, and other formulaic expressions) are expected to be stored as whole units, which helps alleviate the burden that language comprehension poses on short-term memory. Besides these multi-word sequences, research has suggested that a range of frequency- and context-related factors can facilitate comprehension[30] speed. In addition to frequency, predictability plays a vital role in efficient comprehension. Predictability is tightly connected to chunking because the appearance of the first element of the chunk/multi-word expression activates in a chain-reaction type manner the other elements that consequently require less storage space in short-term memory[31]. Based on numerous experiments, van Dyke & Johns (2012; 2006) came to the conclusion that what matters during language comprehension is not the memory capacity of the language users but the stimulus itself (the presence or absence of multiple potential candidates that could be activated and cues for their retrieval) and person-related factors such as their ability to exploit the available retrieval cues. It seems that the efficiency of language comprehension is gradual, just like automaticity itself. Certain linguistic structures can be processed more efficiently than others, and

[30] For the list of these linguistic factors and the empirical evidence see chapter 3.

[31] For empirical evidence that multi-word expressions demonstrate eased processing see Schmitt (2004), for a critical of view of their study consult Siyanova-Chanturia (2013). For a more recent study that empirically validate the processing ease of stored multi-word expression see Yaneva et al. (2017).

language users can demonstrate higher degrees of automaticity than others in certain situations.

5.2.1 Is language comprehension difficult to control?

The question of control is debatable in language comprehension. Once we have mastered the language, comprehension and production mostly happen automatically; without much effort, we decode the written and oral input we are exposed to. One of the most frequently carried-out tasks to validate this claim is the Stroop task. The task instructs participants to suppress reading and focus on colour naming in congruent and incongruent conditions[32]. The experiment's results usually lead to the same conclusion and signal the unintentional nature of comprehension processes. Moreover, it suggests that it is difficult to control or suppress comprehension. The opponents of this claim suggest that language users have developed the mechanism of executive control to deal with interference and confusion in language comprehension. This mechanism helps to regulate and guide cognitive processes in memory. The listener, during comprehension, might rely on this executive control to select among competing interpretations the most appropriate one in the current communicative situation (Ye & Zhou 2009). It is undeniable that skilled language users cannot suppress comprehension in certain communicative situations; thus, it happens in an automatic manner. However, the lack of control during comprehension or production does not mean that this is always the case. When listening to input with poor quality or having a conversation with a non-native speaker with a heavy accent, comprehension becomes a deliberate action where we search for meaning.

Can language comprehension happen without attention and awareness?

The main question researchers ask concerning attention and awareness in language processing is to what extent language users can produce or comprehend language without being aware of it. It must be highlighted that attention does not refer to a single mechanism but to various mechanisms or subsystems, including alertness, orientation, detection within selective attention, facilitation and inhibition. The role of these subsystems is to control information processing when existing skills and routines are inadequate (Schmidt 2010). This suggests that language comprehension can

[32] In congruent conditions, the word and its colour are the same (e.g., the word *green* printed in green ink), while in the incongruent condition, there is a mismatch between the word and its colour (e.g. the word *green* printed in yellow).

demand more or less attention from the listener/reader depending on the communicative situation or the complexity of the input.

According to the Noticing Hypothesis, attention and awareness are the cornerstones of language learning (Schmidt 2010; Robinson 1995). Language processing is necessary for language learning. Thus, the novelty and unexpectedness of the input can trigger conscious attention, while stored memories require less attention because they are expected or emerge as probable in a certain context. Awareness is related to attention; however, the two concepts are not equal. Priming is an excellent example of language comprehension without awareness. Language users are usually unaware that they have been exposed to a prime[33]. The question is whether the priming effect is a good indicator that language processing can happen without awareness. Some researchers adhere to the idea that conscious (aware) and non-conscious processing are simply two stages of perception. During the first non-conscious stage, information is activated automatically, while later, this activated information is consciously processed (Greenwald 1992). This theory was slightly modified with time to become more dynamic. The upgraded theory includes conscious access as a central concept. According to the theory (Dehaene & Changeux 2004; Kouider & Dehaene 2007), the human brain is viewed as a collection of specialised processors that usually operate non-consciously but whose content can be consciously accessed whenever necessary. In other words, there is no clear line between conscious and non-conscious, aware and unaware processing. Rather, language users can attend to one of the several levels of representation at any given time. This view is in line with the idea of the gradual view of automaticity outlined in this dissertation. Language comprehension can happen without control, attention, or awareness if the input is entrenched, but language users have to process it more deeply or have a specific goal during comprehension; they can control comprehension and direct their attention towards that unit that requires it.

[33] Lexical priming refers to faster word recognition latencies following the prior or simultaneous presentation of a meaningfully-related prime word. For example, *night* would be recognized more quickly as a real word in the English language following *day, moon, dark, evening, summer*, or the indirectly related *sun*. In experimental settings, the prime is usually shown to participants for a fraction of a second (50-100 ms.), but it already triggers activation of the unit in the brain. The presentation time is so short that participants are not aware of the fact that they have been shown a word; it appears only as a flash on the screen.

5.3 Automaticity in language production

Automaticity in production can be manifested in both oral and written language. It might seem that automaticity in production is easier to measure as researchers have clear evidence from the output, while during language comprehension, we do not have direct access to the underlying processes of language comprehension but usually rely on reaction times or the saccades of eye movements. The lack of automaticity in production can be signalled by longer pauses or filled pauses, frequent repetitions and restarts of sentences or phrases. Sometimes effortful language production is triggered by a communicative situation like public speaking or writing an official document. For this reason, every researcher knows that there are many hours of direct or indirect practice behind a perfectly delivered conference talk (lecturing and/or considerable experience with giving talks at conferences).

Moreover, language production is a very complex process that includes a number of processes that are not visible to the naked eye or "ear" as they happen prior to the actual production of oral or written speech. These processes include conceptualising our thoughts and selecting the appropriate words and grammatical patterns that help convey our intentions. According to Levelt (1989), these processes are part of executive control and thus require the speaker's or writer's awareness and attention and thus are likely to be controlled rather than automatic. However, not all of these processes are under executive control because adults' experience with speaking is so extensive that certain communicative messages are available as a whole in the long-term memory and can be retrieved holistically (Levelt 1989). For example, prefabricated units fall under this category. Language users, without much thinking, wish each other good morning when they enter the office or express gratitude with the words *thank you* if they are being helped out or given something. In other words, it can be said that communicative messages, depending on their complexity, may require more or less deliberate planning, attention and awareness. This claim emerges from the conceptualisation of automaticity outlined above and supports its gradual nature. Besides conceptualisation and message generating, most of the processes that are part of language production are automatic, especially the selection of the necessary phonemes and their gestural execution (Levelt 1989).

5.3.1 Measuring automaticity in production

In order to measure automaticity in production, just as in the case of comprehension, certain features have to be defined that signal this automaticity. Efficiency and intention can also be regarded as features that determine the degree of automaticity in production (Moors 2016). Efficiency refers to the fact of how easily language production interferes with other cognitively demanding tasks. This efficiency is usually measured when speakers are required to do a secondary task while producing language. These tasks can vary from driving (Kathryn Bock et al. 2007) to walking, finger tapping or ignoring irrelevant sounds (Kemper et al. 2003). Studies like these have demonstrated that language production is not fully automatic and is constrained by capacity limitations to a certain degree.

The second question is whether language production can happen without the speaker's intention. Levelt's (1989) seminal book on language production is entitled "From Intention to Articulation". This title suggests that language production cannot happen without the speaker's clear intention to produce speech. This suggestion would implicate that production cannot be automatic. However, empirical evidence shows that speakers sometimes say something they did not intend or were asked not to. In an experiment, participants were asked not to reveal certain pieces of information during the experiment, but many of them did it unintentionally; this evidence suggests that production can happen under certain circumstances without the speaker's clear intention (Lane et al. 2006).

5.3.2 Other signs of automaticity in language production

In oral production, other signs of automaticity include temporal and phonetic reduction[34]. This particular type of automaticity is referred to as neuromotor automaticity by Bybee (2006). The explanation for this effect is that the articulatory representation of words and sequences of words is made up of neuromotor routines. When sequences of neuromotor routines are repeated, their execution becomes more fluent. This increased fluency results from establishing a new routine, as when a group of words comes to be processed as a single unit. In the new routine, articulatory gestures reduce and overlap as the routine is repeated (Bybee 2006: 714–715).

[34] Literature on temporal and phonetic reduction usually connects these reductive processes to the frequency effect. This dissertation envisages the reduction process as a sign of automaticity in oral production.

Speech-motor control is one of the most complex sensorimotor activities that humans engage in. Producing speech requires a range of processes to happen in a coordinated manner, such as timing and the coordination of muscles that are interlinked. These muscle movements have complex mechanical properties in order to move the diverse articulatory structures of the tongue, lips, jaw, velum and larynx into a wide range of configurations (Parrell et al. 2019).

Linguistic items that are produced often are produced faster and can also show phonetic reduction (deletion, elision, etc.) due to smoother transitions between the diverse articulatory structures. This can be viewed as a different type of automatic process where intention and efficiency play a less vital role. This type of automaticity is clearly a result of practice and exposure intensity. An interesting question to ask is whether this speed of retrieval and articulatory execution is correlated with the strength of the memory representation[35].

"The ease of processing that characterises fluent language production may be the result of knowledge that has been proceduralized through the strengthening of connections, learning sequences of language - from the sequences that make up words to the sequences that make up discourse, or chunking" (Chenoweth & Hayes 2001: 81). A possible measure of fluency in written text could be the words/multi-word expressions/constructions that are produced without a long pause. The lack of a longer pause is likely to indicate that the item is stored as one memory unit and is activated in one step, i.e., it is a holistically stored unit. It is suggested that in written language production, just as in comprehension or oral production, efficiency can be increased by activating chunks, prefabs and other linguistic units that are stored and retrieved holistically. One of the very few studies that have taken up this research was done by Gilquin (2020). In her study, she collected data from non-native speakers of English using special software that time-logged every keystroke[36] the participants made. This method allows measuring writing fluency in the form of bursts of writing[37]. In this particular study, stretches of speech produced without pause in the texts of L1 and L2 speakers were identified as constructions. The results indicate that more automatic production is achieved when participants rely on holistically stored units. Moreover, Gilquin (2020) concluded that not all bursts correspond to constructions, but they

[35]As to my knowledge, no study has looked upon this issue. This question will be answered in the results section of the dissertation.

[36] The name of the software is InputLog and it is freely available here: https://www.inputlog.net/downloads/. For the relevant publication, see Leijten & van Waes (2013).

[37] Burst of writing are defined by Chenoweth & Hayes (2001) as segments of text written between pauses.

might alternatively indicate that mentally stored constructions might deviate from the type of constructions that researchers traditionally identify.

Based on this short description of automaticity, it is apparent that automaticity is undeniably a gradual phenomenon, and the degree of automaticity is subject to situational and speaker-related differences. This has far-reaching consequences for linguistic decision-making processes and the storage of linguistic information. Linguistic elements processed together over extended periods develop into cognitive routines executed automatically; that is, no further decisions are necessary once the routine has been started. In other words, automatisation reduces the number of decisions that language users have to make during language comprehension and processing, which consequently speeds up the whole process (Diessel 2019: 35). However, a process can be automatic despite being controlled or less efficient. The presence or lack of a single feature of automaticity is insufficient to determine whether a process is automatic. It is much more complex because these features' combined effect determines the degree of automaticity. In the following, phonetic reduction as one of the significant manifestations of automaticity in language production will be discussed in detail.

5.4 The nature of phonetic reduction

Language use is a highly automated process involving the fluent and fast production of articulatory gestures, selecting the appropriate words and constructions, and monitoring the addressee. These processes occur in a fraction of a second, and this speedy processing and production would not be possible without automatisation because the production of articulatory gestures is highly automatised (Diessel 2019). This automatised production of articulatory gestures is the main source of phonetic and temporal reduction.

For an extended period of time, reduced forms of words were considered to be "slurred", "slovenly", or otherwise deficient (Zellers et al. 2018). It was claimed that the reduced forms occur solely in casual speech. However, more recent research has shown that reduction is a much more complex phenomenon than was thought, and reduction processes can occur in casual speech and formal situations. Spoken corpora analysis has revealed that language users frequently produce reduced words. Johnson (2004) found that over 60% of the words in the Switchboard (Godfrey et al. 1992) corpora are produced with at least one sound missing in comparison to their full form. Similar results were obtained by Schuppler et al. (2011), who investigated spoken Dutch, and their analysis showed that 40% of the words in the corpus were phonetically reduced. This large number of reduced words that occur in everyday conversations raises interesting questions about the nature of reduction processes, the factors that trigger them, and what role the users' social background and cognitive skills play in these reduction processes. Since speakers produce and listeners comprehend reduced pronunciation variants, the processing of these variants must also be accounted for by psycholinguistic models of speech comprehension and production[38]. The main questions that arise concerning our capacity to produce and comprehend reduced speech include: What are the environments in which reduced forms occur the most frequently? Are reduced words the result of an online process that emerges from the system setting, i.e., the implementation of articulatory gestures in different communicative contexts, or are these forms stored in the mental lexicon? What role do frequency, predictability, and individual differences among language users play in the reduction process? The aim of the discussion that follows is to address these questions and use them to conceptualise reduction as a possible manifestation of automaticity in language production.

5.4.1 Reduction in models of language production

The psychological models that aim to describe language production and processing fall into two distinctive groups that form a continuum. There are abstractionist models on one end of the continuum and, on the other end, the exemplar models (Editorial 2011). The abstractionist models assume that different pronunciation variations are stored in the mental lexicon as abstract representations, such as sequences of phonemes. However, the number of lexically represented forms is limited and sometimes only includes the full forms. In order to explain the reduction processes, these models assume the existence of certain mechanisms, for instance, phonetic implementation rules, that convert full pronunciations into reduced variants during speech. In the exemplar models, at the other end of the continuum, all the tokens ever produced or observed by the speaker are stored in the mental lexicon, including all the acoustic details of these words. According to this theory, the production of the words includes the activation of a particular exemplar representing the word. Some of these exemplars include the reduced forms, which provide an explanation of why

[38] Ernestus & Warner (2011) offer an excellent introduction to reduced pronunciation variants.

speakers are able to produce reduced forms, and why listeners can comprehend them (Editorial 2011).

The middle of the continuum is occupied by the hybrid models, which assume that a word's pronunciation is stored in the mental lexicon as abstract representations and exemplars. In these models, it is especially crucial to determine whether variation resulting from reduction is stored in abstract representations and whether this type of variation has a different status than, for instance, variation resulting from characteristics of the speaker's voice (Editorial 2011).

According to Levelt (1989), among the most prominent characteristics of fast speech are reduction and assimilation. Reduction can arise at different levels of language production. Speakers can achieve reduction by increasing their speed of pronunciation. The speed is increased by generating shorter messages, using a telegraphic register, or retrieving reduced allomorphs. All these processes happen above the level of phonological encoding, i.e., prior to the actual pronunciation. A further possibility of increasing the speech rate is using the weak forms of function words[39]. According to Levelt (1989), selecting weak forms for production is very sensitive to the context in which the words occur, especially to immediately adjacent words. Finally, language users can opt for reducing segments in words. For instance, initial unstressed syllables are reduced to a schwa. This process is likely to happen after the word has already been selected for production. During production, the segments are reduced because unstressed syllables are given such a minimal setting for their duration and loudness parameters that they disappear in articulation.

The way how Levelt (1989) talks about reduction does not allow us to draw any conclusions about the mental representation of the reduced forms. Some strategies to increase speech rate rely on stored representations, such as selecting the reduced allomorphs, while others, such as segment reduction, seem to be an online process. Thus, it seems that Levelt's (1989) theory is somewhere in the middle of the continuum, and it is a hybrid model.

The theoretical framework of this dissertation envisages reduction as a phenomenon that relies both on some stored units in the long-term memory (collocations, valency patterns) and on online processes that emerge from the system setting. In other words,

[39] In a study by Adams et al. (1993) alterations in speaking rate were associated with changes in motor control strategies. In particular, the control strategy for speech gestures produced at fast speaking rates appeared to involve unitary movements that may be predominantly preprogrammed, whereas gestures produced at slow speaking rates consisted of multiple submovements that may be influenced by feedback mechanisms. This preprogrammed movements can be interpreted as entrenched routines.

when speakers activate chunks and other holistically stored units for production, they can show increased speech rate, leading to phonetic as well as temporal reduction. Moreover, if familiar words occur in contexts where they are expected due to frequency or transitional probability, speakers have a ready-made phonological encoding and can execute the articulatory gestures much faster. Furthermore, besides relying on chunks and retrieving words based on their expectedness, reduction processes vary on the dimension of individual differences, such as cognitive skills related to memory and processing speed.

5.4.2 Defining reduction

Reduction is defined by Schäfer (2013) as the non-realisation or deviant realisation of several elements in a word in comparison to a given full(er) form. The same phenomenon is defined by Clopper & Turnbull (2018: 25) as "the phenomenon in which linguistic units (e.g., segments, syllables, or words) are realised with relatively less acoustic-phonetic substance (e.g., shorter duration and/or less extreme articulation) in a given context relative to other contexts". The second definition extends the concept to the dimension of temporal reduction, thus shortening the verbalisation time.

In order to detect phonetic reduction, one has to compare the reduced form with the full(er) form. The fuller form is usually referred to as the careful realisation of the word in clear speech (Schäfer 2013: 3), though there is an ongoing debate about the clear-cut distinction between full(er) and reduced forms[40]. This neat division emerged from the work of Jakobson & Halle (1956). However, recent research claims that this binary division is not sufficient to characterise the different facets of reduction. The first counterargument against this division is that if we assume that reduced forms might be problematic, this assumes that they have a deficit in comparison to the canonical forms (Zellers et al. 2018: 9–10). To a certain degree, this is true in that some phonetic information is missing in the reduced forms. In the Phonetics of Talk-in-Interaction, where "the potential meaningfulness is a basic tenet" (Zellers 2018:10), phonetic upgrades and downgrades describe the difference between a full and a reduced form. The phonetic downgrades lack something that an upgraded form possesses.

[40] Other terms referring to fuller forms include citation form, canonical form or dictionary form; see Zellers et al. (2018).

Lindblom (1990), in his H&H Theory[41], claims that articulation is influenced by constraints on the production system and constraints on the output. In other words, the way a speaker pronounces a certain word depends on their preference for a minimal expenditure of effort and the need to make oneself understood. This means that reduction should not be viewed as a problem case; it is simply a response to a set of system settings. Lindblom's (1990) assumption is that reduced forms lack some information because the language user reduces the pronunciation effort and therefore reduces phonetic information when the listener can understand the communicative message of the language unit without it.

Another problem concerning the binary division is that it implies that one of the forms is more primary. This division suggests that the full(er) form of the words is the primary one. Some researchers, including Abercrombie (1965 cited from Zellers et al. 2018:10), emphasised for a long time that the primary material for linguistic analysis is conversation, i.e., spoken language, which is commonly reduced over the course of communication. The canonical form is given the primary position because word recognition tasks reveal that these forms are recognised and processed faster. One limitation of this finding is that the target words are usually tested in isolation without any context in word recognition tasks, and in such an experimental setting, the canonical forms are more expected than the reduced ones.

According to exemplar theory (Bybee 2002a, 2002b; Pierrehumbert 2001), the distinction between the full(er) and reduced forms is not psychologically valid because language users store more than one pronunciation variation. The use of a particular form is dependent on many factors. One of the most important factors is the context in which the word is going to be used. The more casual and informal the conversation is, the higher the tendency for the reduced form to be used. Another factor that might influence this choice is whether the reduced form can achieve the same communicative goal. If we can indeed achieve the same communicative goal using reduced forms, the chance that the reduced form will be used is higher because it requires less effort from the speaker.

This dissertation seeks to answer whether this neat division between canonical and reduced forms is psychologically valid or whether the mental lexicon is organised in a different manner wherein reduction is seen as an online process and is not bound to stored reduced forms. The assumption about the simultaneous storage of the full and

[41] The theory of Hyper and Hypo-articulation

reduced forms (as proposed by exemplar theory) would imply that during every communicative situation, language users would register the used form and store it together with the communication situation in which it was used, thus whether it was a formal occasion or casual speech. This view, however, would require an enormous amount of memory capacity. It is suggested in this dissertation that rather than viewing canonical (full) and reduced forms as separate forms stored in our mental lexicon, we should think about reduction as an online process that can be triggered by many factors such as frequency, context, and predictability, and it signifies automatic production of the linguistic item. However, this dissertation does not deny by any means that every utterance that is produced or perceived leaves a trace in our memory, as suggested by exemplar theory (Bybee 2002a; Pierrehumbert 2001). When speakers are exposed to reduced forms of words, they are likely to register this occurrence, but it is unlikely they will store every such form as a separate memory trace. Predictable and high-frequency words are more likely to be reduced in all kinds of communicative situations due to the automatisation of neuromotor patterns and due to the fact that reduced forms are more likely to occur in speech. Thus, the approach taken in this dissertation implies that reduction processes are triggered by a range of linguistic factors and facilitated by the language users' cognitive skills, such as processing speed. In other words, reduction does not only depend on the word and its features (such as frequency, predictability, etc.) but also on the speaker who produces it and how quickly the user can retrieve the word from the mental lexicon, and how smoothly they can execute the sequence of articulatory gestures needed to produce the word. The same word can be produced in its full form in one context and reduced in another. Similarly, one person might reduce the word due to familiarity or expectedness, and another would produce the word in full form due to a lack of familiarity despite the fact that the word is expected or is very frequent. This view is in line with the gradual nature of automatisation. In essence, a word can show various degrees of reduction depending on the communicative situation, the context in which the word appears, and the speaker who produces the word. Moreover, phonetic reduction can be gradual and not strictly categorical because of gestural overlaps and the degree of temporal reduction is determined by how smooth the articulatory gestures are.

5.5 Measuring reduction in language production

The degree of phonetic reduction can be assessed using segment and word duration measures, vowel space expansion, and f0, among others. The discussion of reduction in this chapter will be limited to phonetic variation along measurable acoustic dimensions, namely temporal (i.e., duration) and phonetic/segmental (i.e., deletion of phonemes) reduction.

5.5.1 Phonetic reduction

There are three processes that can generally be regarded as phonetic reduction: deletion, lenition, and assimilation. Deletion[42] at the segmental level can be defined as "the absence of a segment that is present in the citation form" (Schäfer 2013: 4). Studies that deal with this phenomenon include the study of schwa and /t/ deletion in German by Kohler (1991) and final t/d deletion in English by Bybee (2001). Lenition refers to "the weakening of an articulatory gesture in the pronunciation of a certain sound" (Schäfer 2013: 5). This articulatory weakening is different for different classes of sounds and therefore has different acoustic correlates. If one compares the processes of deletion and lenition, the differences are the following: deletion is a categorical process while lenition is gradual, which means that a speech sound is either deleted or not, but the weakening of a speech sound can be partial and happen in a stepwise manner. The processes seem to have clear-cut features, but it is sometimes debatable which reduction type characterises the phenomenon under question better. Firstly, deletion is often referred to as a categorical phenomenon, but it can be gradual in some instances. Thus, sounds may be very short and weakly articulated but still be present. The t/d deletion can sometimes be seen as gradual, resulting in a gestural overlap and not a complete deletion. Secondly, deletion does not always indicate the complete absence of a speech sound (Schäfer 2013). For example, Bybee's (2001) study about Spanish in New Mexico showed that the deletion of the nasal sound resulted in the nasalisation of the preceding vowel. Researchers agree that lenition and deletion are reductive processes, but assimilation is controversial in this context. Assimilation can be defined as the process of "one sound taking over characteristics

[42] The term elision is also used to refer to the same process.

of a neighbouring sound with regard to place or manner of articulation." (Schäfer 2013: 5-6)[43].

5.5.2 Temporal reduction

Temporal reduction is a less straightforward concept. The commonality between phonetic and temporal reduction is that to define them, we need a basis, i.e., a canonical form in the case of phonetic reduction and a clear[44] speech pronunciation in the case of temporal reduction. These two reduction processes (phonetic and temporal) are tightly interwoven. Reduced vowels and consonants will automatically lead to shorter word duration, and speedy pronunciation can trigger co-articulatory processes resulting in different phonetic reduction processes such as assimilation or lenition. The theoretical framework of this dissertation claims that temporal reduction is the result of the word or word combination's automatisation in production. Automatisation is the result of repetition, and the degree of automaticity is mainly determined by external factors such as frequency, probability, familiarity, context, and the language users' cognitive skills, such as memory capacity and processing speed. When discussing second-language speakers, temporal reduction can be defined in terms of fluency in some instances. Segalowitz (2010: 6) suggests that fluency is a behaviour that

> …reflects the execution of the neurological and muscular mechanisms that a speaker has developed over an extended period of time through socially contextualised communicative activities. The operation of mechanisms reflects the cognitive and emotional states of the speaker at the time of current state of the speaker.

This definition can be applied to explain the neuromotor automaticity that is observed in temporal reduction. Temporal reduction can be observed in words and word combinations, which are produced so often that the neurons responsible for the execution of single phonemes are synchronised to the degree that their activation is automatic. The linguistic unit is accessed through the lexical route; that is, the sublexical orthographic information makes direct contact with the whole-word orthographic representations, which then provide access to whole-word phonology on the one hand, and higher-level semantic information on the other. This assumption raises the question of how we should define the role of frequency in temporal

[43] cf. Herbst (2010: 73)

[44] By clear speech the careful realisation of the phonemes is meant. Clear speech includes less reduction and usually has longer pronunciation time Clopper & Turnbull (2018).

reduction: does frequency refer only to how often the word is observed in the input, or does it also incorporate production intensity? In other words, can frequent input without extensive production lead to the development of neuromotor automaticity? Hochberg (1976) argues that fluent reading is always mediated by the creation of a speech articulation program[45]. Additionally, Rayner (2006) reports that reading experiments with various paradigms have established that skilled readers activate phonological information when reading silently. These authors suggest that even when we are just listening to or reading the input, we strengthen the neuromotor routine responsible for the linguistic item's smooth production.

5.5.3 Phonetic reduction in German

The experiments within the framework of this dissertation focus on the German language. That is why, in the following, the view on phonetic reduction in German will be discussed. The discussion will be limited only to elision of intervocalic /t/ and pre-nasal schwa. Research on reduction processes in German most often treats reduction as something that emerges from the phonetic environment. The section has the objective of highlighting in which phonetic environment the reduction of these sounds occurs. The discussion focuses on the rules determining the phonetic context facilitating reduction processes.

One of the most prominent researchers in the field of phonetic reduction in German is Kohler. In his book *Einführung in die Phonetik des Deutsches* (Kohler 1977), he draws attention to the fact that words in isolation and sentences lead to different articulatory movements. The claim supports Clopper & Turnbull's (2018) theory on "easy" and "hard" contexts that are claimed to define the rate of reduction processes. Moreover, it can be explained as a result of the fact that words have a co-articulatory influence on each other. He emphasises that any kind of reduction processes that happen due to coarticulation are not random, though it may seem so; on the contrary, they are determined by strict physiological, articulatory, linguistic, and situational conditions. The physiological and articulatory trigger of these changes is the fact that the different articulatory organs differ with respect to precision and speed of carrying out articulations; firstly, because of a differing degree of muscular activity; secondly, because of the changing nature and number of back referring sensors (Kohler 1977: 208). His explanation of phonetic reduction is completely rule-based, assuming that

[45] also see Haber & Haber (1982)

reduction would always happen if the condition were met. The approach taken in this dissertation is that reduction happens when the word becomes automatised and occurs in a context that triggers the automatic production of the item. It is not denied by any means that this reduction can happen only in a particular phonetic environment, but the presence of the environment is not enough to trigger reduction. The most frequent reduction processes in German include (Kohler 1977, 1991; O'Brien & Fagan 2016):

a) elision;
b) geminate reduction;
c) weak forms.

These reduction processes are claimed to be regulated by several different rules that determine the contexts in which reduction is possible or unlikely to happen. In the following, these rules will be briefly described. The most frequently elided sounds in German include the /ə/ the /t/. The /ə/-elision is the result of the stressed–time structure of the language, where pitch changes tend to happen between vowels. This phenomenon leads to the deletion of the /ə/ in a pre-nasal position. In this case, the nasal overtakes the vowel's function when it is after an apical word-final consonant or in front of another consonant; in front of the vowels, the number of syllables is reduced. The following rules constrain the rate of /ə/ elision (O'Brien & Fagan 2016; Kohler 1977):

a) schwa cannot be deleted if the following consonant is an obstruent;
b) schwa is usually deleted before /m/, /n/, /l/, /R/;
c) the most restrictive environment for schwa deletion is when /m/ follows schwa since the sound that precedes schwa must be a fricative;
d) schwa deletion before /n/ is allowed only if the sound preceding schwa is an obstruent (plosive or fricative)[46];
e) when /l/ follows schwa, the sound preceding schwa must be a consonant;
f) there is no restriction on the sound that precedes schwa if /R/ follows;
g) the word-final schwa is omitted in verbs in the first person singular present tense when there is a pronoun in proclitic or enclitic position, but only if no syllabic nasals appear because of the deletion.

A second frequent deletion process in German is /t/-elision. The rules of /t/-elision are the following:

a) if the /t/ occurs in the middle of a cluster of three consonants, it is deleted;

[46] There is an exception to schwa deletion before /n/. Although an obstruent precedes schwa in the diminutive suffix {-chen}, schwa is not deleted.

b) after an /n/ or /l/ before an /s/ as a result of co-articulatory shift: e.g., Gla*nz* [nts] → [ns], erhäl*tst* [lts] → [ls];
c) before an /l/ after fricatives, especially after homorganic /s/, e.g. *restlich* [stl] → [sl], *schriftlich* [ftl] → [fl];
d) when assimilation rules are applicable: (a) apical plosives and nasals before word and morpheme borders can be assimilated to the following labials or velars, e.g., *anbinden* [nb] → [mb], *die angebundenen Pferde* [ndnnp] → [mbmmp]; (b) the assimilation also occurs when the consonant cluster is the result of /ə/-elision, e.g. *buntem* [ntm] → [mpm]), e.g. *Weltkugel* [ltk] → [lk];

The above-mentioned rules clearly define the phonological contexts where reduction can happen. No other factors are mentioned, such as the frequency of the words, their probability in the context, or anything that refers to the fact that repetition can also foster phonetic reduction. These rules are necessary to understand which words and sounds are likely to undergo reduction, but they are not enough to discuss automaticity as a manifestation of entrenchment. Moreover, most researchers consider reduction to be a characteristic feature of causal speech; thus, if the scope of factors that trigger reduction processes is widened to include social factors, it takes into account the formality of the context but neglects any kind of possible variation due to the cognitive skills of speakers. Within this dissertation, it is claimed that reduction can occur even in a somewhat formal context, like an academic lecture or read-speech. In other words, phonological properties of words are the results of highly practised behaviours associated with the vocal tracts of human beings (Bybee & Hopper 2001). These highly practised behaviours can occur in casual communicative situations, just as in formal ones. Speech production consists of sequences of different articulatory positions that speech organs make while pronouncing a certain word. With repetition, the transition between the speech organs' movements becomes smoother, and any additional effort not required to pronounce the word is reduced. This results in a shorter word voicing duration or the reduction of certain sounds (Bybee 2001). These transitions between the different articulatory gestures indicate the neuromotor automaticity that was developed due to practice. Practice is undoubtedly connected to frequency because it is more likely that people will produce high-frequency words more often than low-frequency words. The phonetic context, without a doubt, contributes to the development of smoother transitions between articulatory gestures, but it is unlikely that it is the *only* triggering factor.

5.6 Conclusion on entrenchment as a product

By way of conclusion, it should be stressed that various factors associated with automaticity do not necessarily co-occur, and so particular linguistic structures should be considered as *more* or *less* automatic (i.e., more or less entrenched) rather than *fully* automatic or controlled, emphasising the gradable character of the phenomenon. Automaticity is present both in language production and comprehension. It has different manifestations, including efficient processing, smooth comprehension and production, and temporal and phonetic reduction. These manifestations can co-occur, but the presence or absence of a single feature is not sufficient enough to determine whether the process is automatic or not. Moreover, different language users can rely more or less on these manifestations of automaticity depending on their cognitive skills or the situational context.

Furthermore, it is difficult to give a unified definition of automaticity due to a large number of linguistic, social, and cognitive factors that trigger it and the different forms it can take. If we aim to operationalise automaticity, it is fruitful to use different methods and participants with various social and cognitive backgrounds. The connection between language users' cognitive skills, language history and their capacity to automatise different processes can shed new light on the whole process.

6 Literature review

This chapter aims to review the literature on entrenchment. It is divided into three main parts. First, the studies that focused on phonetic and temporal reduction will be reviewed. As next, studies that define entrenchment as fluent processing will be discussed. Finally, empirical studies focusing on familiarity as a direct measure of the degree of entrenchment will be highlighted.

6.1 Studies on phonetic reduction in language production and perception

Research on phonetic reduction has a long history. Some linguists, including Jakobson & Halle (1956), proposed that reduced forms are not worth studying as they are lacking in some way. However, not all researchers held the same view, and a number of phonetic studies have investigated topics related to reduction. At the time of Jakobson & Halle's writing, of particular interest was lexical stress. With time, the study of phonetic reduction moved beyond analyses of word stress, and researchers became interested in the contexts in which reduction occurs (Zellers et al. 2018). This was the time when Kohler (1977) formulated the rules that determine schwa elision in German. Moreover, reduction was analysed from a sociolinguistic perspective. Phonetic reduction in the study of Labov (1972) was seen as an expression of group membership rather than as a set of nonprocedural, individual phonetic cases. Later on, reduction was investigated from two different perspectives: from the perspective of motor economy (Lindblom 1990; Kohler 1990), a process that is triggered by certain linguistic features (such as frequency), and the context in which the word occurs (Bybee 2002b, 2001, 2007a; Pierrehumbert 2001, 2002; Jurafsky 1996). As spoken corpora became available, this research area boomed because the correlation between pronunciation speed, reduction and different frequency measures could be directly investigated. The vast majority of these subsequent studies focused on phonetic reduction in language production. The unified observation in previous work on phonetic reduction processes has revealed that there are certain contexts where reduction is more likely to happen. These contexts facilitate the processing and production of linguistic items. The linguistic and contextual factors that create these contexts and trigger phonological reduction include:

a) *word frequency* (Bybee 2002b, 2001; Pierrehumbert 2002; Gahl & Garnsey 2004; Tremblay et al. 2011; Brown & Raymond 2012; Balota, D. & J. Chumbley 1985; Mousikou & Rastle 2015; Pluymaekers et al. 2005);

b) *bigram frequency* (Brown & Raymond 2012; Arnon & Priva 2013);
c) *predictability* ((Jurafsky 1996; Jurafsky et al. 2001a; Jurafsky 2003; Aylett & Turk 2006; Gahl & Garnsey 2004; Tremblay et al. 2011);
d) *segmental factor and syllable type* (Jurafsky et al. 2001a);
e) *speech rate* (Jurafsky et al. 2001a; Gahl 2008);
f) *syntactic category and orthography* (Gahl 2008);
g) *phonological environment* (Bybee 2002a; Kohler 1977);
h) *neighbourhood* (lexical, phonetical, orthographical) (Munson & Solomon 2004; Gahl et al. 2012).

The effect of lexical frequency on phonetic reduction processes has been observed in the temporal domain for words and single vowels (Arnon & Priva 2013; Aylett & Turk 2006; Bell et al. 2009a; Gahl et al. 2012) in the spectral domain for vowel reduction (Munson & Solomon 2004; Bybee 2002a), in continuous speech production (Arnon & Priva 2013; Aylett & Turk 2006; Bell et al. 2009a; Gahl et al. 2012), and in the production of isolated words (Munson 2007; Munson & Solomon 2004; Myers & Li 2009). The right phonological context seems to increase the effect of frequency further. Bybee (2002b) argues that a word with a preceding favouring environment will have more exemplars with deletion than the words that do not have such a favouring preceding environment. In English, final /t/ and /d/ deletion has an alternating environment both in preceding and following contexts. This means that the preceding context alternates, just as the following word context alternates. Bybee's (2002b) hypothesis for final /t/ and /d/ deletion in New Mexican Spanish was that words that occur more often before vowels would exhibit less deletion. This hypothesis assumes that the deletion is no longer purely conditioned by the phonetic environment but that the effects of the change are represented lexically (Bybee 2002b: 275). This hypothesis was supported by a study where the Switchboard corpus's entries were analysed. The study showed that deletion rates for contracted negotiation ('*nt*) were approximately the same for both pre-vocalic and pre-consonantal environments. The conclusion of the study is that

> "as lexicalization proceeds and the exemplar cluster changes, the overall distribution of the exemplars will reflect their distribution in the experience of the speaker at first, and eventually the more frequent exemplars will take over in more and more contexts" (Bybee 2002b: 276).

The analysis of speech samples from the Switchboard and LA corpora showed that it is not simply the word's frequency that leads to phonological changes but the frequency of the word in the context that conditions the sound change. The results of the study by Brown & Raymond (2012) support this claim. They have found that preceding and following favourable contexts contributed positively to the reduction process, just as the predictability of the preceding and following phones.

In some studies, the effect of frequency was not as straightforward as the previous research suggested. Cohen et al. (1990) have not found any significant frequency effect on the durational shortening of English homophones. The frequency inheritance theory (Jescheniak & Levelt 1994) claims that a given phonological form is activated whenever a word associated with that form is activated. This would mean that items that share the same phonological form should behave as though they have the same frequency as well. A study carried out by Jescheniak and Levelt (1994) showed that low-frequency words with high-frequency homophones were produced as fast as high-frequency words. This assumes that the phonological representation is the locus of frequency information in the mental lexicon. Almost a decade later, Caramazza et al. (2001) designed a study similar to the one conducted by Jescheniak and Levelt. They failed to replicate the previous study's results and concluded that lemma representation is not distinct from form representation and homophone pairs have entirely independent representations without a shared phonological form. Gahl (2008) also analysed the duration of English homophones. The results of the investigation revealed that homophone pairs in a corpus of spontaneous speech differed in duration, with high-frequency words being shorter than their low-frequency homophones. Based on these results, she came to the conclusion that form frequency, or the frequency of particular combinations of segments, is insufficient for predicting which forms shorten. Instead, lemma frequency, that is, frequency indexed by information about a word's meaning and syntactic properties, is a determinant of word duration. According to her, this implies that the shortening of frequent forms is not purely the result of increased motor fluency.

In some other studies, n-gram frequency was the primary factor facilitating reduction processes because of growing evidence that speakers are sensitive to the properties of multi-word sequences and draw on such information in production, comprehension, and learning (Arnon & Clark 2011; Frank & Bod 2011; Tremblay et al. 2011; Gradoville 2017). Tremblay et al. (2011) explored the effect of frequency, probability, and mutual information on four-word sequences' production. The amount of experience the participants had with an N-gram (frequency) surfaced as the most critical predictor family in the analysis of production duration. In another study, Arnon & Priva (2013) found multi-word frequency effects on phonetic duration in adults. In their study, they used continuous speech elicited from the Switchboard corpora. The results showed that the frequency of the phrase was the main predictor of the pronunciation length, supporting the findings of Tremblay et al. (2011). They concluded that the phonetic durations are affected not only by features of the word itself or the words appearing directly after and before it but also by the properties of the larger linguistic context.

Of the contextual factors, semantic predictability is the most frequently investigated. It captures a range of different phenomena connected to the semantic, syntactic and extralinguistic context in which linguistic items are produced. High-frequency function words tested for phonological reduction by Jurafsky et al. (2001a) did not show a tendency towards reduction despite their high frequency. This finding initiated extensive research on phonetic reduction as a consequence of predictability. The assumption that predictability leads to production variability resulted in the "Probabilistic Reduction Hypothesis: word forms are reduced when they have a higher probability" (Jurafsky et al. 2001a: 229). According to Jurafsky et al. (2001a), the probability of a word occurring in language use depends on many factors, including the context in which the word occurs, lexical structure, semantic expectations, and even some discursive factors. To support their claim, they analysed 38,000 phonetically hand-transcribed words from the American English telephone conversations corpus. This study made a distinction between the conditional probability of the target word given the preceding word and the conditional probability of the target word given the following word[47]. In order to count the probability of the word, Jurafsky et al. (2001a) used the relative frequency of the word in the corpus and counted the probability of the word occurring in the context and the probability that a string of words occur together in the context. The results showed that the previous word's predictability contributes to the reduction process; the higher the contextual predictability of a word or phoneme, the greater the likelihood of vowel reduction or durational shortening. These shortening effects provide strong evidence for the existence of probabilistic links between words in the mind of language users (Jurafsky et al. 2001a). Another study that focused on the effects of predictability on the rate of reduction was carried out by Aylett & Turk (2006). They investigated the correlation between syllable duration and language redundancy. They defined redundancy as differences in the predictability between elements in a given speech string. They

[47] These two measures later became to be known forward and backward transitional probability.

proposed an inverse relationship between syllable duration and predictability arising from lexical, syntactic, semantic, and pragmatic factors. The rationale behind this claim is that it provides an efficient way of ensuring that elements with low levels of language redundancy are produced longer and with more salient acoustic characteristics. Their results have shown that redundancy factors predict between 20% and 65% of the variation in syllable duration in non-phrase-final monosyllabic words. Gahl & Garnsey (2004) analysed how durational shortening is influenced by the predictability that arises from the different syntactic structures in which the words are likely to occur. The probability used in their study was based on the different sentence structures certain verbs evoke. The main focus was on the duration of the verb in different sentence structures and the duration of the item in selected noun phrases. They concluded that the probabilities on the grammar level could predict the selected words' duration. Thus, differences in pronunciation reflect the probability of a given syntactic structure in a given context. In essence, this variation reflects contextual probabilities of syntactic structures and not just information about which words are likely to be adjacent to each other. The primary result of the study was that contextual predictability of a syntactic pattern, given a particular verb, affected the pronunciation of words in those structures. The distribution of /t, d/ deletion and the duration measurements indicated that verbs were more likely to undergo /t, d/ deletion in high-probability syntactic contexts than in low-probability contexts. Moreover, clause boundaries that had low probability affected verb and pause durations to a greater extent than those with a high probability.

Discourse mention is another contextual factor indicating whether the target word is new or old. Repeated words are expected to be easier to produce and comprehend. Repetition or second mention was a primary factor in many studies (Kahn & Arnold 2015; Curl et al. 2006; Baker & Bradlow 2009).

A handful of studies have investigated how these different linguistic and contextual factors interact and, in this way, influence the rate of phonetic reduction. Baker & Bradlow (2009) examine the combined effect of lexical frequency, previous mention and speaking style on word duration. The results revealed that high-frequency words exhibit more second-mention reduction than low-frequency words in plain speech but not in clear speech. Word durations in clear speech were significantly longer than durations in plain speech, indicating that clear speech involves hyper-articulation[48]. In

[48] Hyper-articulation is part of Lindblom (1990) theory of hyper- and hypo-articulation. Hyper-articulation involves pronouncing words more clearly than they are normally pronounced and is associated with

conclusion, speakers reduce the most when all factors support reduction in plain speech, yet there appears to be a limit on word duration reduction in clear speech. This work has opened up a new avenue of research examining the interactions between different linguistic and contextual factors. However, so far, not so many studies have investigated this research area[49].

Relatively few studies have investigated reduction in German. Those studies that investigated the phenomenon mainly focused on the temporal reduction of vowels and consonants. A study by Mooshammer & Geng (2008) deals with acoustic and articulatory manifestations of vowel reduction in German. They talk about stress-induced vowel reduction as the possible consequence of three processes: (a) duration shortening, (b) more extensive coarticulation, and (c) saving of articulatory effort. Of course, it is possible that all these processes simultaneously contribute to the reduction of vowels. In the study, they collected spoken data from seven subjects. To make some general conclusions and reduce the influence of speaker-dependent differences, they normalised their data with the help of generalized Procrustes analysis. The research questions of the study were the following: (1) is vowel reduction in German (phonological neutralisation is not a particular feature of the language) context-dependent, i.e. the result of the coarticulation with the neighbouring sounds or the centralisation of vowels; (2) Are tense and lax vowels reduced the same way and to the same degree; and (3) Are there any similarities or differences between patterns of acoustic and articulatory target undershoot with respect to their vowel space areas and directions of reduction?

The subjects of the study were seven native speakers of German aged between 25-40. They were recorded by means of electromagnetic midsagittal articulography[50]. The target words likely to exhibit phonetic reduction were analysed as /tVt/ sequences with the full vowels in stressed and unstressed positions (Mooshammer & Gang 2008:122). Each target sound cluster was embedded in a phrase. The study results concerning durational changes show that aspiration duration was shortened significantly in the stressed items compared with all others; temporal duration did not

various acoustic-phonetic features of enhanced speaker effort, such as longer duration. Hypo-articulation involves pronouncing words less clearly than normal, and can involve features such as shorter durations, reduced vowel space and deleted phonemes.

[49] This will be one of the main research questions of the studies carried out within the framework of this dissertation.

[50] Most of the studies mentioned so far were based on spoken corpora. Using corpora makes data collection and analysis much easier, however, it does not allow for a detailed analysis of interspeaker variation.

show any significant differences between stressed and unstressed lax vowels. Unstressed tense vowels showed a significant durational reduction. There was no clear tendency concerning the durational shortening of lax vowels in an unstressed position. The study's articulatory data showed that vowel reduction in the consonantal context influenced both the low vowels, which were elevated and the back vowels, which were fronted. The whole vowel space was moved towards the articulatory place where the following consonant was produced. This leads to the conclusion that unstressed vowels are, to a greater extent, influenced by surrounding consonants.

The study concluded that vowel reduction in German follows the pattern of target undershoots found in other languages (Mooshammer & Gang 2008:131). These target undershoots result from an extensive coarticulation of neighbouring elements. This conclusion is similar to the one that Kohler (1990).

Studies focusing on English have followed a usage-based approach; for this reason, the frequency of occurrence and the context in which the words were used were given a prominent role. Zimmerer & Reetz (2011) analysed German phonetic reduction from a generative point of view. In the study, they compared the "reduced" schwas and the underlying schwas. Their study shows that schwas that result from the reduction process differ from the underlying schwas in some respect. The differences were small and not identical for female and male speakers of German. Materials for the study were elicited from the Kiel corpus of spontaneous speech. They analysed 53433 vowels produced in the corpus and concluded that 3685 vowels are underlying schwas, and 432 (1%) are reduced to schwas. Based on these results, Zimmerman and Henning claim that reduction of vowels to schwa is not common in conversational German, but it occurs (2011: 2325).

The comparison of the studies carried out in English and German regarding phonetic reduction shows that in English, psycholinguistic factors such as frequency, predictability, and the strength of the exemplar are taken into account. On the other hand, in the German research field, rule-based and biological constraints seem to be the basis of reduction. The fact that phonetic reduction in German is not investigated from a psycholinguistic perspective provides the chance to gain new insight into the processes, constraints, and factors that facilitate phonetic reduction and shortening of verbalisation time.

As emerges from the overview of the research on phonetic reduction, frequency plays a crucial role in determining the rate of phonetic reduction and a tendency to shorten the duration of phonation. As frequency is an inevitable component of entrenchment, logically, it plays a role at all stages of entrenchment, including the learning stage, memory encoding, and automatisation. Most of the studies focused on language production and how frequency and other language-related measures determine the reduction rate or influence the selected items' pronunciation time. One of the rare studies that analysed the perception of reduced words was carried out by Lorenz & Tizón-Couto (2019). This study was interested in investigating the role of chunking and procedural strengthening in the comprehension of phonetically reduced forms. Moreover, they asked the question of how frequency and conditional probability influence chunking and procedure strengthening. Procedure strengthening can be understood from their description as the automaticity of optimisation. What is unique about this study is that it measured comprehension and how reduced forms affect information processing. Their target items included a list of *V-to-V*$_{inf}$ constructions, where the particle *to* was reduced in some cases and presented in its full form in others. The two main variables they analysed were the participants' reaction time, i.e., how long it took to recognise whether the particle was present in the input or not, and accuracy. The factors that were used in the analysis to predict reaction times and accuracy included:

a) condition (full or reduced form of the particle);
b) frequency of the construction;
c) forward and backward transitional probability;
d) verb duration, syllable count and verb form;
e) merged plosive cluster;
f) preceding sound;
g) gender and age;
h) control before the target and item count.

Their results revealed that conditional probability played an important role in recognising the reduced forms. There was a positive correlation between age and reaction times. The participants produced faster reaction times to *to* when the preceding verb had a longer duration, regardless of the condition. A longer verb provided a longer processing span, prompting a faster recognition of the upcoming item. Finally, when the final segment of the verb preceding *to* was an alveolar plosive (e.g., *need to*, *hate to*), the cluster ([dt] or [tt]) gets merged into a single segment – this case inhibits recognition of a reduced *to* but does not affect full forms (Lorenz & Tizón-Couto 2019: 768). These results suggest that if the input is expected, it facilitates the recognition of the reduced form, i.e., when the quality of the input is poor, this stored knowledge compensates for it when the phonetic cues are weaker. Moreover, they suggest that chunking, just as entrenchment and automaticity, should be viewed as a continuous process, i.e., more entrenched and chunked units are likely to be stored and retrieved as one single memory. Less entrenched and less chunked units can be computed online during language processing and production.

Although extensive research has been carried out on phonological reduction, no studies (to the best of my knowledge) have extended the scope of linguistic and context-related features to possible social and cognitive predictors. In most of the reported studies, phonological reduction processes were seen as properties of words. The frequency, context and predictability of the word facilitate the reduction processes. This would assume that reduction is almost sure to happen if the right features are present. In other words, reduction is seen as a property of words. Within the framework of this dissertation, reduction is seen as the product of automaticity, which is subject to individual differences. This suggests that words in the right context can demonstrate different reduction rates depending on the speaker's social background and cognitive capacities. Analysing the interaction between linguistic, social, and cognitive factors can deepen our understanding of the reduction processes and the factors that trigger them.

6.2 Studies on processing speed in language perception as a manifestation of entrenchment

Besides phonetic reduction, another field of research that can be connected to entrenchment includes studies that analyse language processing and how certain linguistic and cognitive factors determine the ease and fluency of language processing. One of the only studies that directly aimed to measure entrenchment was carried out by Blumenthal-Dramé (2012). Moreover, this study analysed entrenchment not only from the perspective of psycholinguistics but also from a neurolinguistic perspective. The process of chunking and holistic storage was the basis of her study. The stimuli for her study included suffixed bimorphemic derivatives like *kissable, employer* or *careerist*. The neurolinguistic experiment was a primed lexical decision task. The aim of the study was to find out to what extent pure token frequency obtained from corpora can predict entrenchment. The study hypothesised that higher entrenchment should make complex strings more unit-like in the sense that their processing would resemble monomorphemic words of equal letter length. The prime for the study always included the root of the bimorphemic derivative (e.g., *cheer* for *cheerful*). The results of the behavioural data showed that contrary to usage-based expectations, the token frequency of the derivatives was not predictive of entrenchment. Relative frequency turned out to be the most significant predictor[51]. Another interesting outcome of the study was that the size of the phonographic[52] neighbourhood was negatively correlated with the degree of entrenchment. The neurolinguistic study examined the degree of BOLD[53] activation in high and low relative frequency conditions. The results confirmed the findings of the behavioural study and concluded that the mental connection strength between two stored entries (derivative "whole" and its monomorphemic constituent "part") depends on the surface and base frequency, the letter ratio between the whole and its part, and the number of neighbours (Blumenthal-Dramé 2012: 184). This was one of the first studies that questioned the role of frequency in entrenchment and that it can be seen as a direct measure of entrenchment. This study pushed forward the experimental investigation of other language-related metrics and their role in the process of entrenchment.

In two related studies, McConnell & Blumenthal-Dramé (2019, 2021) investigated a) how the task and corpus-derived association affected the online processing of collocations and b) the role of usage-based individual differences in the probabilistic processing of multi-word sequences. The first study aimed to empirically investigate the psychological validity of a range of corpus-based metrics such as log-likelihood, dice coefficient, and z-score. Bigram frequency and forward and backward transitional probability of collocations were also taken into account. Online processing was measured in the form of a moving window self-paced reading[54] test. The target collocations were embedded into sentences, and reaction time was the main

[51] Relative frequency in the study of Blumenthal-Dramé (2012) is composite measure which results from dividing surface by base frequency, and which has also been claimed to affect morphological processing. Surface frequency refers to raw frequency of the bimorphemic word (e.g., *government*) and the base frequency is the frequency of the base of the bimorphemic word, i.e. *govern*.

[52] Phonographic neighbours refer to the number of words which can be obtained by changing a single element (phoneme, letter, or both)

[53] Blood-oxygen-level-dependent imaging, or BOLD-contrast imaging, is a method used in functional magnetic resonance imaging (fMRI) to observe different areas of the brain or other organs, which are found to be active at any given time.

[54] This particular type of self-paced reading is called moving-window word-by-word self-paced reading. The moving-window part refers to the fact that only one word is shown at a time and that pressing SPACE both advances to the next word and hides the previous word.

measure of processing ease[55]. The statistical modelling results have revealed that bigram frequency and backward transitional probability were the most significant predictors of the participants' reaction times. Interestingly, none of the corpus-based association metrics reached the significance level, suggesting that these metrics are not cognitively realistic. The significance of bigram frequency supports the idea that frequently co-occurring items form chunks and are holistically stored in our mental lexicon. In the second study, individual differences were added to measure how usage history, including reading experience and age, can predict the online processing of multi-word units. The methodology was the same as in their previous study. This study included a range of tests that aimed to highlight how individual differences determine the processing speed of certain modifier-noun combinations. These tasks included a vocabulary test, an author recognition task and a 'words that go together' task[56]. The author recognition test aims to define the participants' reading habits quantitatively. This study confirmed the results of the previous one and showed that backward transitional probability is a significant predictor of processing ease. Age was negatively correlated with the reaction times, i.e., older participants processed the target items faster than younger participants. This suggests that they rely more on predictive processes, and this knowledge is more entrenched in the mind of older participants than in those of the younger ones. Their suggestion is based on the assumption that this stored probabilistic knowledge could be the result of greater experience with language and general world knowledge. Surprisingly none of the measures quantifying individual differences reached the significance threshold.

As emerges from the results of the studies that investigated entrenchment in language comprehension, there is a need to explore the issue in more detail. None of the aforementioned studies investigated the interaction of the different language and cognition-related measures. The central claim of the theoretical framework of the current dissertation is that the cumulative effect of these factors can shed new light on the concept of entrenchment and the underlying processes that lead towards entrenchment and define the degree of entrenchment.

[55] The collocations all consisted of an adjective and a noun. Reaction time was extracted for each noun and was used as the dependent variable for the mixed-effects regression models. Due to the high correlation among the linguistic predictors, they were analysed separately.

[56] The test was first introduced by Dąbrowska (2015c). It focuses on the collocational knowledge of the participant. Later the task was extensively analysed by Garibyan (forthcoming) in a crosslinguistic study including German, Russian and English.

6.3 Studies on the role of familiarity in the empirical validation of entrenchment

A handful of studies have measured the degree of entrenchment in the form of familiarity scores. Familiarity is related to frequency, and frequency without question is one of the most often-measured predictors of entrenchment. Familiarity can be defined as a speaker-related measure of entrenchment. Familiarity scores are expected to represent all speakers' cumulative usage and exposure frequency. These studies take advantage of different sociolects (e.g., the language of lawyers, engineers, and doctors) because the language users' experience with the item can help demonstrate how frequency varies according to different registers depending on the social and educational background of language users.

One of the first studies that questioned the predictive power of corpus-based frequency measures was carried out by Gardner et al. (1987). Their population consisted of law students, engineering students, and nurses, and their methodology included a simple lexical decision task. The stimuli were divided into six categories: engineering, medical, low-frequency nontechnical, medium-frequency nontechnical, and two groups of high-frequency nontechnical words. Their main aim was to demonstrate that part of the frequency effect is determined by the participants' experience with the lexical units and not the general frequency of the words in the language. Moreover, they aimed to show that low-frequency words can trigger fast reaction times if the participants are familiar with those words due to their social background. Their results have shown that reaction times and accuracy were correlated with word frequency and the social background of the participants, i.e., nurses reacted faster to medical terminology and made fewer errors during the lexical decision task. The study results highlighted systematic differences in lexical decision times due to subjects' occupational backgrounds and the technical categories from which words were drawn. This was one of the first studies that drew attention to the fact that frequency should not be regarded as a universal quantifier of experience with language because something that has low frequency for an average speaker might occur very frequently in other speakers' language use.

A more recent study has explicitly connected familiarity to the mental state of entrenchment (Caldwell-Harris et al. 2012). Their study investigated the properties of frequent multi-word expressions. They proposed that the mental representations of language users have formed a continuum, from word combinations that have fossilized into single units (*nightclub*) to those that both exist as independent units and yet have bonds, varying in closeness, with the words with which they frequently co-occur. Multiword expressions of varying lengths can also become entrenched and obtain status as units (Caldwell-Harris et al. 2012: 166). The stimuli used for the study included adjective + noun and noun + noun pairs. The frequency of these pairs was obtained from the COCA[57] corpus and from Google.

Moreover, they also obtained familiarity ratings for each of the multi-word expressions. The methodology they used to determine mental entrenchment was a perceptual identification task[58]. The participants consist of Orthodox Jews and secular Jews. Respectively, the multi-word expressions they used were phrases from religious rituals with different patterns of use across groups with different prayer habits. The participants' familiarity with the multi-word expression was measured using a sentence completion task and a detailed questionnaire about their praying habits. The results have revealed that phrases from daily prayers had higher accuracy than phrases selected from weekly and annual prayers, and these frequency effects were stronger in religious participants. The explanation they provided is that annual phrases were more entrenched than would be expected by a once-yearly recitation because of their high emotional charge. Thus, they directly correlated the results of the perceptual identification task with the entrenched status of the phrases. This suggests that individuals with greater experience with specific linguistic expressions had greater accuracy at reading the briefly displayed phrases, indicating stronger memory representations.

One of the most recent and comprehensive studies on the connection between familiarity and entrenchment was carried out by Verhagen (2020). In a series of studies, she investigated variations in metalinguistic judgments about multi-word sequence processing between and within participants. One of the studies aimed to discover the extent to which familiarity judgments are related to usage frequency and are influenced by the context. The participants of the studies included recruiters, job seekers, and people not (yet) looking for a job. The reason for this selection was motivated by the fact that these groups can be expected to differ in their experience with word sequences that typically occur in job ads but are not expected to differ systematically in their experience with word sequences characteristic of news reports.

[57] COCA - Corpus of Contemporary American English. Available here: https://www.english-corpora.org/coca/

[58] In a standard perceptual identification task, a stimulus is briefly displayed on a computer screen, typically for durations of 30 to 100ms., and then masked with a visual noise pattern, which disrupts continued processing.

Moreover, she tested the reliability of familiarity judgements and their correlation to corpus-driven frequency measures in her series of studies. The familiarity judgement of the participants was largely influenced by two factors: 1) the extralinguistic context in which the language users encounter the stimulus, and 2) the extralinguistic context the word strings evoke, i.e., the frame in which the stimulus is activated and assessed. The results of the first study that aimed to validate the reliability of the familiarity judgements have shown that corpus-driven frequency measures were the main predictors of the familiarity judgement scores. Embedding the selected phrases in sentences did not affect the familiarity ratings. Concerning the stability of the judgements, participants provided very stable patterns of familiarity ratings as a group. However, once the rating stability within each participant was measured, the picture became less stable, and large differences between the first and second ratings were observed. The follow-up study asked the question of whether the rating process (Magnitude estimation vs Likert scale[59]) influenced inter-and intra-individual variation across time. An analysis of variation across time showed that there was no significant difference between the two rating systems.

The study that aimed to underscore the connection between one's language experience and the mental status of the selected stimuli included both sentence completion and a voice onset task. The stimuli included phrases frequently used in job advertisements and phrases that frequently occur in news reports. The results of this study showed that there are significant differences between groups and item types; this was especially visible when comparing the performance of the three groups to the job ad stimuli. The recruiters demonstrated the highest proportion of correct responses in the sentence completion task in comparison to the other two groups. This difference was not observed among groups when analysing the news report phrases. This result can be explained by the fact that there is a strong correlation between previous language experience and the participants' expectations about upcoming linguistic elements. The voice onset time task results confirmed these results; inexperienced participants responded faster to the news report stimuli than to the job ad stimuli.

Moreover, the results revealed that there was a direct correlation between mentioning the word in the sentence completion task and voice onset time. Interestingly, the sentence completion task results significantly influenced the participants' voice onset

[59] During Magnitude estimation task participants are asked to judge the familiarity of stimulus in comparison of the previous stimulus, i.e., is it more familiar they should give a higher score, if it less familiar then a smaller score respectively. When the Likert scale is used, they always have to choose a number between 1-5 (in some other versions between 1-7).

time more than the surprisal score generated by language models. This once again supports the idea that generic corpus data cannot adequately assess the mental lexicon of the participants. Furthermore, Verhagen (2020) concluded, based on her series of studies, that the correlation between familiarity ratings and the participants' performance indicated the degree of entrenchment in their long-term memory. This assumption leads to the conclusion that the degree of entrenchment is part of the participants' metalinguistic knowledge.

7 Experimental design

7.1 Study description

This section describes the experiments that aim to empirically validate entrenchment. The four experiments and the case study presented in this section will examine the relationship between usage frequency, familiarity, reading habits, corpus-driven association measurements and entrenchment. The studies will focus on different linguistic levels; the first two experiments focused on the level of words, and the last will concentrate on adjective-noun word combinations embedded in sentences. The experiments will make use of various psycholinguistic methods, including rapid word naming, self-paced reading, and verbatim recall, to provide evidence for entrenchment.

The experiments exploring entrenchment on a word level will mainly highlight the effect of frequency and word familiarity on the temporal and phonetic reduction rate. The follow-up study widens this scope and includes transitional probability and dispersion as additional factors contributing to the investigation of the linguistic predictors that facilitate the above-mentioned reduction processes.

Overarchingly, the role of individual differences in speakers is investigated, with a focus on working memory capacity and differences in cognitive capacities and styles. The rationale behind examining two different levels, i.e., words and adjective-noun combinations, is that variables that correlate with significant facilitation both on the level of words and word combinations must be considered predictors of entrenchment. Furthermore, the main experiment will focus on automaticity and the strength of memory representations. It will be aimed across all the experiments to establish a connection among the predictors, processes, and the product of entrenchment. Automaticity is operationalised within the framework of this dissertation as the product of entrenchment. Automaticity can take many forms; in all four experiments, temporal and phonetic reductions will be the focus of the investigation. These two reduction processes are treated as manifestations of automaticity in language production. The experiments gradually extend the scale, including the cognitive effects and psychological affordances[60]. Because the first two experiments are less comprehensive in scope, they cannot be regarded as cornerstones of validation or falsification of entrenchment but can rather be considered as

[60] Schmid (2017a) refers to the effects entrenchment has on the cognitive system as cognitive effects and psychological affordances.

supplementary experiments to deepen our understanding of the phenomenon of entrenchment and to support the development of a test battery that can capture the full complexity of the phenomenon.

The description of each experiment will be introduced with an outline of the data collection and analysis. The studies will then be described by the study design, participants and the equipment used to carry out the study, followed by the results and discussion. The studies and the development of the research methods are summarised in Figure 4.

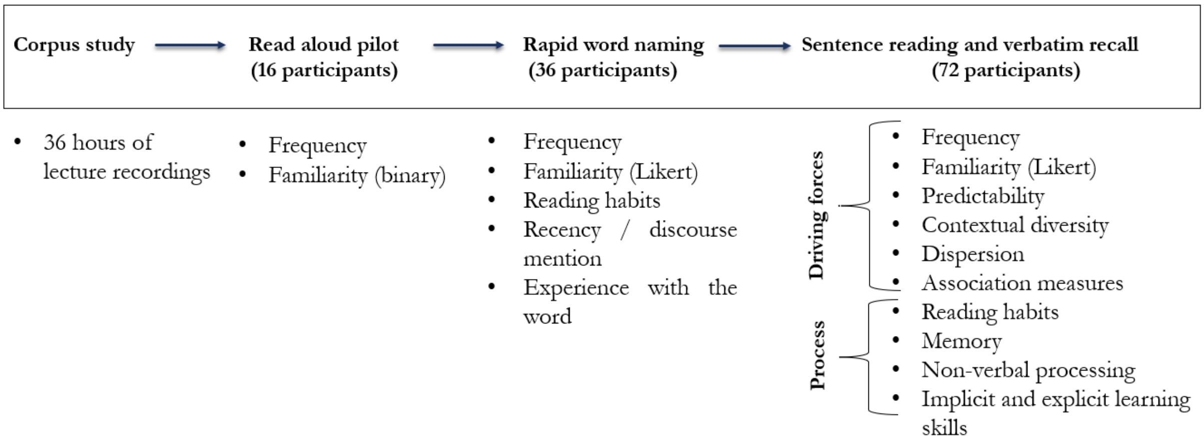


Figure 4. Summary of the experiments

7.2 Corpus study

The corpus study aims to investigate the extent to which recency and frequency can determine the rate of temporal reduction and the type of frequency (frequency of exposure vs production) that can better predict the rate of temporal reduction. This study has an exploratory rather than a hypothesis-driven goal. Temporal reduction is treated as one of the linguistic manifestations of entrenchment. The study will only focus on this specific linguistic manifestation of entrenchment. Bybee's exemplar theory (2002a, 2002b) attributes a special role to temporal reduction caused by the frequent production of certain words or word combinations. This type of temporal reduction is referred to as neuromotor automaticity in the works of Bybee. Neuromotor automaticity can be defined as the automatic execution of articulatory gestures resulting from frequent production. This study explores how the rate of temporal reduction changes within one lecture and across fifteen lectures.

7.2.1 Data collection

The materials for the present corpus study come from a law lecture series held by a single lecturer over the course of 4 months. The video recordings of the lectures were collected from the online video portal of the University Erlangen-Nuremberg with the permission of the lecturer. The course and identity of the lecturer will be kept anonymous. The field of law is rich in terminology, which is used mainly by those who work or study in this field. In the course of investigating the impact of frequency on temporal reduction, it is important to have as much information as possible about the frequency of exposure and the production of the target words. Using recorded lectures provides numerous advantages in examining the role of recency and frequency in great detail: one can measure how frequently the lecturer produced the target word during one lecture as well as over the course of the semester and compare this to the frequency of the words in a law corpus. This provides the possibility to analyse the role of repetition and offers an effective way of explaining the connection between frequency and neuromotor automaticity, where the frequency of production is a crucial factor. The majority of previous studies on the effect of frequency on phonetic and temporal reduction focused on English, more precisely on American English, using the data obtained from the Switchboard Corpus (Godfrey et al. 1992). In an effort to extend the research base, the present study will investigate the phenomenon of temporal reduction in German.

The data collection was done with the help of the Regional Computer Centre Erlangen. The video recordings of the lectures were processed with the European Media Laboratory's (EML) natural language processing software. The recogniser uses a time-synchronous beam search algorithm that utilises a pronunciation lexicon organised as a prefix tree for the dynamic creation of the search space (Fischer & Kunzmann 2013). The final output files contain a text document and an XML file. The text documents were compiled into a corpus with the help of Sketch Engine (Kilgarriff et al. 2014). The resulting corpus contains 272420 tokens from approximately 36 hours of recorded natural language. From the compiled corpus, a wordlist with frequency information was created. 48 nouns were selected from the corpus and later reduced to 25 to suit the main criteria of the target words occurring in the majority of the lectures. The frequency of the nouns varies from 785 to 29. All the target words are specific to the field of law.

7.2.2 Data analysis

The first step in analysing the role of frequency in temporal reduction was to determine the length of pronunciation of each occurrence of the selected words in the lectures. This information was extracted from the XML files[61]. The extracted pronunciation lengths were summarised in a data frame that included

a) the frequency of uttering the word within a lecture,
b) the overall frequency of the word in all the lectures,
c) the frequency of the word in the BGH-Urteile Corpus[62].

As mentioned above, the study aims to determine whether the more frequently pronounced words will correlate with temporal reduction, demonstrating automatic production. One way to examine this phenomenon is to track how the length of pronunciation changes within one lecture and during the 4 months. However, if one wants to investigate whether the selected words are uttered more quickly not only due to the effect of repetition but also because they are entrenched in the mental lexicon of the speaker, one has to compare the speech rate during the pronunciation of the selected words with the average speech rate of the lecturer. A higher speech rate can indicate temporal reduction due to neuromotor automaticity.

Calculating the average speech rate is more complicated than it seems at first glance [63]. The literature introduces several different measurements. The most common among them are words per minute (wpm), syllables per minute (s/min), syllables per second (syll/sec or s/s), average syllable duration (ASD in msec), phones per second, or average phone duration (in ms). All of the above-mentioned units of measurement have different advantages and disadvantages. Measuring the speech rate in words per minute is one of the easiest ways, but at the same time, it is the most unreliable. Both mono- and polysyllabic words occur within a sentence or text, and the distribution of these words will determine the speech rate. Utterances that include more

[61] The script, which parsed the XML files and extracted the words, is based on the ElementTree package of Python.

[62] The BGH-Urteile Corpus is a collection of decisions of the Federal Court of Justice that contains 50431804 words. It was compiled by Stefan Evert and his team at the Friedrich-Alexander-University Erlangen-Nürnberg.

[63] As Trouvain (2004) indicates, speed is relating a distance covered by a body to the time used. If we relate it to speech production, then the body refers to the articulatory organs. Some of the organs make a greater number of movements than the others, for example, the movements of the tongue compared to the movements of the velum. Some of the speech organs do not move at all, for example, the palate or the upper teeth. That is why measuring distance in millimetres or other units of measuring distance is not applicable to the field of linguistics. Trouvain (2004: 42–43) offers the following explanation to the fact why other units of measurements have to be used to measure speech rate: "speech is the result of the temporally coordinated execution of articulatory gestures that lead to speech events."

monosyllabic words might result in a higher speech rate than utterances that include a higher number of polysyllabic words but fewer words overall. Another problem that arises in this context is determining exactly what a word is. This includes the challenging question of hyphenated words and whether they should be considered as single or multiple words.

One of the most frequently used measurements in the literature is the syllable per second. Syllables are units above the individual sound segments. When talking about syllables, one must bear in mind the distinction between the so-called underlying syllables derived from the words' lexical form and the realised syllables in speech. It is a matter of deciding which definition of the syllable is more relevant for measuring speech rate for each particular study. The discrepancy between the number of realised and underlying syllables can indicate temporal or phonetic reduction. In this study, the number of underlying syllables will be taken as the basis of speech rate measuring. The same problem is encountered when measuring speech rate in sounds per second. The IPA transcription of a word reflects the number of sounds that make up the word and is an impractical measure. In speech production, the sounds undergo phonological changes due to many factors, including the speaker's articulation and the communicative situation. Functional words tend to be shortened in connected speech, and most of the vowels are reduced to shwa. The vowels in unstressed syllables of the lexical word go through the same process (for a more elaborated discussion of the advantages and disadvantages of the speech rate measuring units, see Trouvain 2004). The average syllable duration seems to be the most appropriate measure to investigate temporal reduction. Trouvain (2004) claims that the quality of the vowel usually determines the length of the syllable, and because the peak of the syllable is a vowel, the average duration of the syllable can indicate the speaker's speech rate.

Adopting any of the methods mentioned above to measure speech rate brings about technical problems[64]. De Jong & Wempe (2009)automatised the speech rate measurement with the help of the PRAAT (Boersma & Weenink 2018) software. Their

[64] To divide the number of syllables pronounced during the lecture by the duration of the lecture is not a manually performable task. As mentioned previously, the fifteen lectures correspond approximately to 36 hours of speech. Even if the syllables were calculated automatically and would be divided by 36 hours, this would give a misleading speech rate. First, because the lecturer pauses a couple of times, the length of the pause varies depending on the situation. He might be waiting for an answer to a posed question or looking for a note. Secondly, in the .txt file, it is not visible when the lecturer is speaking and when, for example, a student is answering a question.

script aims to determine the syllable's nucleus and, with the help of this information, counts the syllables and then determines the speech rate[65].

The speech rate of the lecturer was determined separately for each lecture. The script mentioned above was used to calculate the speech rate in PRAAT and counts two types of speech rates. The first is calculated by dividing the number of syllables by the duration of the lecture. The second one, called articulation rate, is calculated without the lecturer's pauses, which is the phonation time divided by the number of syllables pronounced during the lecture. To inspect whether the calculations were correct, the same process was done with the audio files containing only single words. The word *Beweisverwertungsverbot* (prohibition of the use of evidence) was analysed to verify the script's accuracy. In the fifteen lectures, this word occurs 318 times[66].

The result of the analysis of the word *Beweisverwertungsverbot* showed that, according to the script, the word has a different number of syllables throughout the lecture varying from four up to eight. The word has seven underlying syllables. Due to the different syllable counts, the speech rate of the pronunciation of this word is slower than the average speech rate.

Of course, this raises the question of whether the reason for this syllable count is the fact that the word underwent phonological reduction or whether the software fails to detect the number of syllables correctly due to the fast speech rate. Upon investigation, it was found that all the syllables are pronounced in the audio recording but with a faster speech rate. The differences between the peaks and the dips are not audible and might explain why the software fails to determine the exact speech rate.

In addition to the PRAAT script, a Python-based script was implemented to count the number of syllables in the lectures. The difference between the two syllable-counting

---

65 The script executes the following steps in order to count the speech rate: 1) it extracts the intensity (higher energy) with the parameter of 50 Hz as the minimum pitch using autocorrelation; 2) it considers all the peaks above the threshold of 0 to 2 DB (decibel) as possible syllables; 3) it inspects the preceding dip in intensity and considers only a peak with a preceding dip of at least 2 or 4 dB concerning the current peak as a potential syllable; 4) it extracts the speech contour using a window size of 100 msec and 20 msec time steps and excludes all the peaks that are unvoiced;
5) the remaining peaks are saved and are considered as syllables. The script is freely available online: *sites.google .com/site/speechrate*.

66 PRAAT displays the results in a so-called info window where the following data are presented: the name of the audio file, the number of the syllables, the number of pauses, duration of the recording, phonation time, speech rate and phonation rate and the average syllable duration. The software also determines the number of pauses. The default setting of the script regards 300 milliseconds of unverbalised speech to be a pause. The definition and the length of a pause is a debatable question in psycholinguistics. While some studies work with a threshold below 50 ms (pause and segment detection with an automatic procedure), in other (mainly psycholinguistic) studies, only silent pauses longer than 500 ms are regarded (De Jong & Rutger Bosker (2013). Other examples are 200 ms Grosjean & Collins (1979); 150 ms Tsao & Weismer (1997); 130 ms (Dankovičová (1997); 100 ms de Pijper & Sanderman (1994).

methods resulted in a difference of 7468 syllables. This discrepancy between the number of syllables accounts for a statistically significant difference (p-value = 0.01855) in the speech rate 4.454 syll/sec versus 4.264 syll/sec. Both PRAAT and the Pyphen script-based speech rates were used in further analysis. In addition to the speech rate, the phonation rate was also included to avoid long pauses unduly influencing the speech rate. The speech rates were calculated separately for each selected word and lecture and summarised in a data frame.

To compare the frequency measurements mentioned above, a mixed-effects regression model was used to analyse the data (RStudio (2015), packages lme4 and lmerTest), where the length of pronunciation was treated as the dependent variable and the selected words as random effects. Frequency information from the law corpus was log-transformed before fitting it to the model[67]. A decrease in the length of pronunciation is treated as temporal reduction, which is seen as the linguistic manifestation of entrenchment.

First, a global model was fit to the data with all the explanatory variables (frequency in a lecture, overall frequency, length of the word, lecture order, and log-frequency in BGH-Urteile). The result showed that three out of five explanatory variables, namely lecture order, frequency in BGH-Urteile and word length, predict the temporal reduction rate. Contrary to expected, the overall frequency of the words in the lectures has no significant influence on the rate of temporal reduction. The first global model is summarised in Table 3.

Table 3. The estimated coefficients, t-values and p-values for the mixed-effects regression model fitted to pronunciation lengths across the lectures

Predictor	**Estimate**	***t* Value**	***p***
Intercept	5.084e-01	4.834	<.001
FreqLect	-3.142e-	-0.328	0.74
Lecture 9	-3.452e-02	-2.664	<.01
Lecture 12	-3.014e-02	-2.154	<.05
Lecture 13	-3.230e-02	-2.365	<.05
Lecture 14	-4.558e-02	-3.389	<.001
Lecture 15	-2.677e-02	-1.712	.08
Occurrence	-9.573e-05	-0.982	0.32
log(Freq_BHG)	-2.383e-02	-2.665	<.05

[67] The analysis was done dollowing the suggestions of Baayen (2008) and Levshina (2015)

WordLength	3.471e-02	9.926	<.001

In the second step, three models were fitted to the data. All of these models followed the following template (PronLength ~ SIGNIFICANT VARIABLE + (1 | WORD)). The three models were then compared by analysing variance (ANOVA, R version 4.1.0). In line with our expectations, the model with word length as an explanatory variable was significantly better than those including the log-transformed frequency information or lecture. A baseline model was fitted to the length of pronunciation times to see the role of the frequency-based explanatory variables. This model contained only word length as a fixed effect and the words as a random effect. The rationale behind fitting such a baseline model was to correct the response variable for effects known to modulate pronunciation length but independent on the experimental manipulation (Linzen & Jaeger 2015). The per-word residualised pronunciation lengths of the baseline model are the corrected pronunciation lengths that were used for further analysis.

The newly fitted model shows that the effect of frequency disappeared, and only the order of the lectures (coded as a factor from 1 to 15) significantly affected the temporal reduction rate. A possible explanation of why the effect of frequency disappears when the word length is statistically controlled for is the interaction between word length and frequency and that high-frequency words tend to be shorter.

The assumption behind treating lectures as an explanatory variable was to see if there is a steady increase in the rate of temporal reduction towards the end of the term. The figure below shows that the length of pronunciation was stable during the fifteen lectures, but lectures 6,7 and 13 had longer pronunciation times. Due to this variation, the preceding lectures turned out to be significant predictors of temporal reduction in the mixed-effects regression models. This variation, however, is more likely to be explained by some extralinguistic factors (mood of the lecturer, students' participation) than by the effect of repetition. The variation across the lectures is represented in Figure 5.

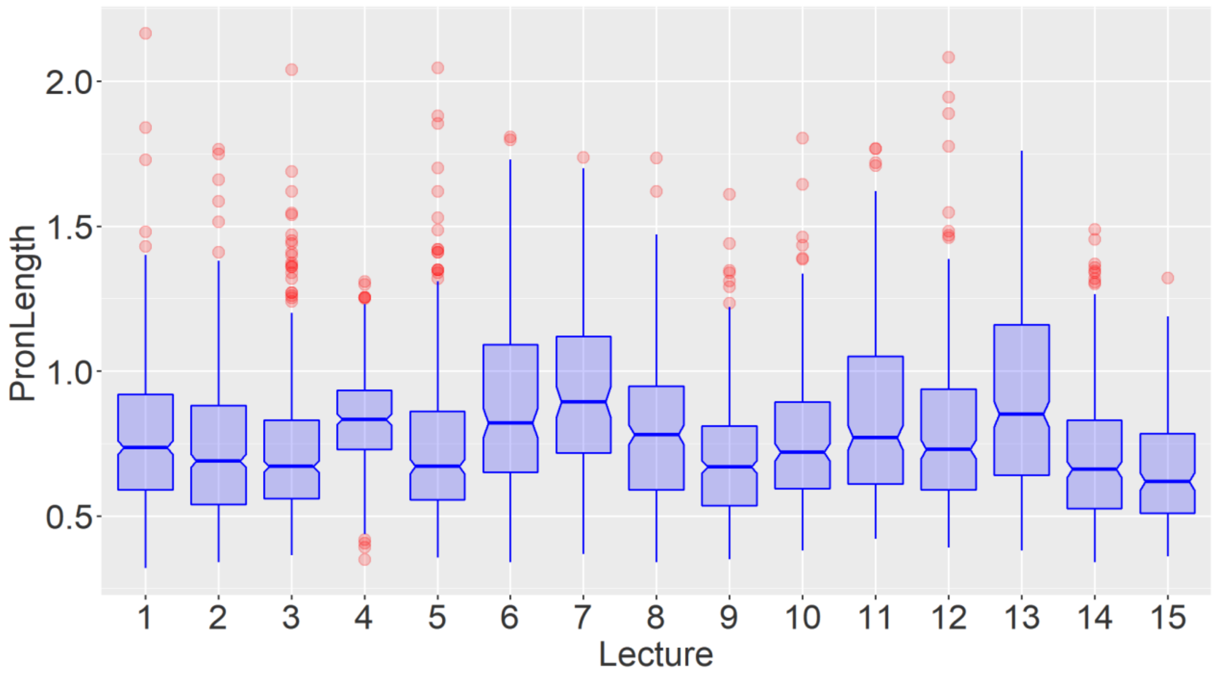


Figure 5. Variation in pronunciation length across the lectures. Y-axis represents the pronunciation length of the selected words in seconds; the x-axis represents the number of lectures.

A factor that did not significantly affect temporal reduction was repetition, measured as the frequency of production during the single lectures and the whole term. A new dependent variable was also calculated to measure speed up, i.e., which words triggered higher pronunciation speed than the average pronunciation speed during the lecture. For this, the pronunciation rate of the item was divided by the average pronunciation rate during the lecture.

The only difference between the models was that using pronunciation speed as a dependent variable led to higher[68] R^2 (marginal R^2/conditional R^2 0.542 / 0.622 vs marginal R^2/conditional R^2 0.110 / 0.371).

A one-tailed t-test was used to compare the differences between the speech rates to determine which measures should be used in the following experiments. The rationale behind the comparison was to determine whether the different measures were equally reliable and whether the lecturer's speech rate was higher during the pronunciation of the selected words compared to the average speech rate. The speech rate of the selected words was compared first to the automatically measured

[68] R-squared (R2) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model.

speech rate using the PRAAT script (t = 76.143, df = 4860.9, p-value < 2.2e-16), then to Python-based script measurement (t = 67.873, df = 4866.9, p-value < 2.2e-16), and finally to the articulation rate (t = 25.672, df = 4834.9, p-value < 2.2e-16). In all three cases, the difference between the speech rates is significant. The lecturer has a higher speech rate while pronouncing the selected words in comparison to his average speech rate.

The differences between the measurements are illustrated. Figure 6 shows that the speech rate calculated using the PRAAT script is the slowest (LectP), and the articulation rate (speaking time divided by the number of syllables) is the highest, which is expected because, during a lecture, longer pauses are natural.

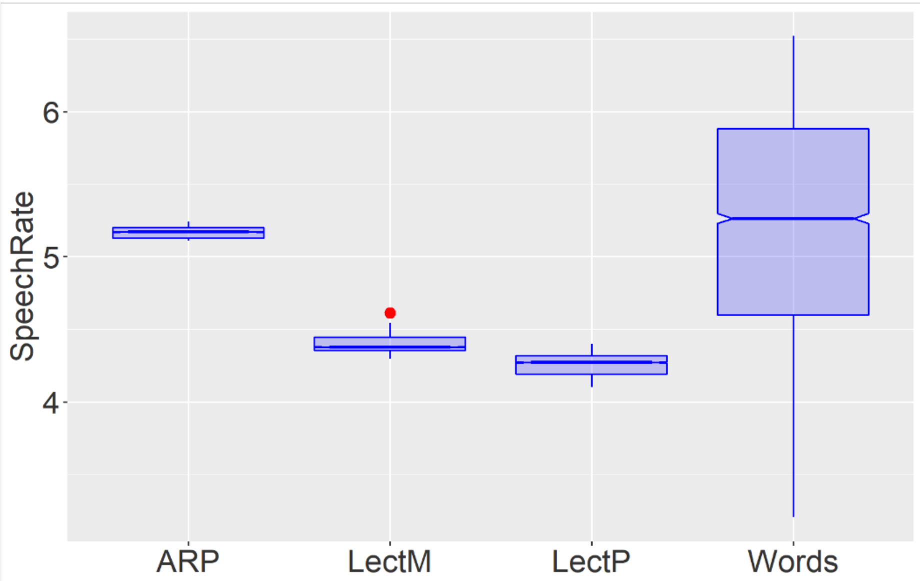


Figure 6. Differences between speech rates and speech rate measurements. ARP – articulation rate during the lectures using the PRAAT script, LectM – speech rate measured using the Python script, LectP – speech rate calculated using the PRAAT script, Words – speech rate during the pronunciation of the selected words.

7.3 Behavioural studies

This section will introduce the design for each of the four experimental conditions used in this dissertation. Two out of the four experiments (experiments 1 and 3) are considered pilot studies due to their limited scope. Each section will first discuss the aim of the study, research questions, and hypothesis, followed by the stimulus description. The different tasks that were part of the four experiments will be discussed in one section to avoid repetition. In the interest of readability, predictor variables such as FREQUENCY will henceforth be printed in SMALL CAPS.

7.3.1 Experiment I (pilot study)

The study aimed to explore the extent to which familiarity can predict the rate of temporal reduction. Two groups of participants were recruited; the first group comprised law students, and the second, participants outside of the field of law. It is assumed that the participants from the area of law will show a faster and more effortless production of the target words due to the higher frequency of production and increased exposure. This assumption is based on the case study results, where a higher speech rate was observed during the pronunciation of the selected words in comparison to the average speech rate of the lecturer.

7.3.2 Experiment II

This study aimed to evaluate and validate the role of word familiarity that was observed in the previous study and build upon this observation with a more detailed familiarity judgement test. The role of reading in the development of neuromotor automaticity (manifested in the form of temporal reduction) was additionally investigated. The study asks the following research questions:

1) To what extent do differences in the amount of experience with a particular register manifest in the rate of temporal reduction? (Law vs non-law participants)
2) To what extent do participants' responses in the familiarity judgement task predict their production speed?
3) To what extent can the frequency of exposure (measured in the form of reading habits and corpus-elicited frequency measures) determine temporal reduction?
4) To what extent do the familiarity judgements correlate with the frequency of the words?

7.3.3 Experiment III (pilot) + Experiment IV[69]

The results of the first two studies revealed limitations. First, using single words made it possible to compare the speech rate of the participants directly, but on the other hand, it made the experiment more artificial because words are typically embedded in sentences in natural language. The second limitation was found in the source of the frequency information, which should be extracted from more than one corpus. Third, frequency alone might not be enough to explain the variation across the items. The final limitation is that if we want to understand the complex character of entrenchment, we have to explore factors beyond speech fluency.

The final experiment attempts to implement the knowledge gained from the previous studies. The test battery is more complex than that of the first two experiments in that it is complemented by tasks measuring other cognitive abilities of participants, e.g., memory span and processing speed, implicit and explicit learning ability, and measures of reading habits and word familiarity. Instead of using single words, participants are exposed to whole sentences that include the target items. This test battery more closely resembles natural language than the previous ones. In addition to the frequency of the words, the context and transitional probability of the words are being controlled for.

It was aimed to test most of the linguistic and cognitive effects that entrenchment manifests[70], including fluency in speech, chunking, representational strength, automatic processing, and single-step memory retrieval. The study aims to provide answers to the following research questions

1) To what extent is the degree of entrenchment part of a speaker's knowledge? Furthermore, can we use this knowledge to gain a deeper understanding of the concept of entrenchment?
2) To what extent can the cognitive and linguistic effects predict the degree of entrenchment?
3) How strongly and directly do prior knowledge and recent exposure to the item influence the rate of temporal reduction?
4) How precisely do metalinguistic judgements and processing/production measurements reflect the degree of entrenchment of the target linguistic items?

[69] Experiment III and IV follow a very similar structure and ask the same research questions. Based on the results of the pilot study (experiment III), however, the test battery was slightly changed.

[70] For a detailed list of factors associated with entrenchment see Schmid (2017a)

5) To what extent does the inter and intra-individual variation depend on the cognitive abilities of the language users?
6) How much variation is to be expected among adult native speakers who belong to the same group? Moreover, how can the variability that is observed be analysed?
7) To what extent do familiarity judgments correlate with the corpus frequency of the words?
8) To what extent does the chunking status of the items predict fluency of production, phonological reduction rate, and memory representation strength?

The research questions mentioned above are used to test the following hypotheses regarding linguistic representations and metalinguistic knowledge of the participants and to answer a still-open question concerning the role of individual variation in the process of entrenchment:

Hypothesis I Variation in temporal and phonetic reduction across items correlates with linguistic measures that focus on frequency.

Hypothesis II Variation in temporal and phonetic reduction across participants correlates with the participant's familiarity, reading habits, and verbal/nonverbal processing speed.

Hypothesis III The participants' processing speed correlates with their speech rate, while the working memory span of the participants predicts the accuracy of their sentence recall.

Hypothesis IV Nouns preceded by high transitional probability adjectives are accessed and pronounced more quickly and recalled more accurately.

Hypothesis V In the delayed recall condition, only the items with strong memory representation are recalled correctly.

Hypothesis VI Participants substitute the low-transitional adjectives with the high-transitional ones after being exposed to them.

7.4 Stimuli

Experiments I and II focus on nouns (bi- and multimorphemic), and experiments III and IV focus on adjective-noun combinations that fall into high-frequency and high-transitional probability. For each experiment, except experiment I, the stimuli are neatly divided into items belonging to the sphere of law and those to be found in everyday language.

7.4.1 Experiment I

The 12 nouns of varying frequency selected from the law corpora[71] were embedded into a text. The only criterion that was used during the stimuli selection was that the words should vary in their frequency both in the general and law corpora, and they should be law-specific terminology. The text resembles a newspaper article about a court case[72]. The usage of the words was analysed in two leading German journals - *Spiegel* and *Zeit* to ensure the idiomaticity of the text. According to Bybee (2001), the context contributes to the activation of the word. It was attempted to use the words with their typical collocates. Native speakers of German edited the text to avoid grammatical mistakes and ensure that the text sounded idiomatic. The story told in the text is not based on any real events. The names of the people and places are invented to avoid connections between real events and the event described in the text. The stimuli and their frequencies are illustrated in Table 4 below.

Table 4. Stimuli used for Experiment I

Word	Frequency DWDS	Frequency Lectures
Beschuldigtenvernehmung	1	3
Beweisverwertungsverbot	1	263
Bundesverfassungsgericht	866	48
Ermittlungsmaß1hmen	5	35
Ermittlungsverfahren	203	260
Gerichtsverfassungsgesetz	100	10
Klageerzwingungsverfahren	8	4
Rechtsmittelverfahren	10	6
Staatsanwaltschaft	1540	505
Strafverfolgungsbehörden	46	80
Verfahrensverzögerung	1	7
Zeugnisverweigerungsrecht	52	85

[71] The corpus was compiled based on the 36 hours of lectures and it included 272420 tokens.

[72] The text that was used for experiment one is to be found in the appendix.

7.4.2 Experiment II

The words for experiment II were selected from the compiled lecture corpus. 30 nouns that belong to the field of law and occur at least 30 times in the lecture corpus were included in the experiment. The words vary in their frequency; the most frequent word occurs 780 times, and the least frequent one occurs 30 times. An equal number of distractor words were chosen. Each target word was matched with a distractor word of the same letter length. According to lexical access theories, the size of the word measured in terms of its length in letters affects the recognition times (Warren 2013). To find equally long and equally frequent distracts is almost impossible. The Zipfian law claims that the length of the word is in reverse correlation with its frequency. Some of the target words are 25 letters long. German is rich in compounds, which made it possible to find words with the same length but not always equally frequent. All the distractors occur at least seven times in the DWDS corpus. The list of words is included in the appendix.

7.4.3 Experiment III

In the initial step, a list of critical bigrams representing everyday language and combinations frequently used in the field of law were elicited from two corpora. The general words were extracted from the German Web corpus using the [tag = "ADJ.*"] [tag = "N*."] CQL query. The legal bigrams were elicited from the BGH-Urteile corpus using [pos="ADJA"] [pos="N.*"] query. All the bigrams were modifier noun sequences obeying the following constraints: the modifier was either an attributive adjective or a (past or present) participle, and modifiers were limited to certain nouns with which they can co-occur. Additionally, it was aimed to select nouns that consisted of one lexical morpheme. All the nouns belonging to the general category satisfied the last requirement. It was more difficult to find legal nouns that consisted of one single morpheme. 80 bigrams were selected that complied with the above-mentioned criteria. Next, the nouns were removed from the bigrams, and the list with the modifiers was distributed among 100 students to test whether the elicited bigrams have a similar representation in the mind of language users as in the corpus, i.e., whether participants will use the same nouns as those extracted from the corpora. Participants were instructed to write down the first noun that came to mind after reading the given adjective. Additionally, they were provided with examples to make the task clear. The list of adjectives that belong to everyday language was distributed among linguistics students who were all native speakers of German (35 students). The adjectives that belong to the field of law were given to law students (65 students).

The adjective-noun combinations provided by the students significantly resembled those obtained from the corpus, although the ranking of the bigrams based on their likelihood to occur together is slightly different. The corpus-elicited bigrams were ranked according to their transitional probability. The equation provided by Jurafsky et al. (2001a) was used to calculate it. The equation is as follows: $P(w_i) = \frac{C(wi)}{\sum_i C(wj)} = \frac{C(wi)}{N}$.

Two different types of probabilities are distinguished; the forward and backward transitional probability. The forward probability is also referred to as prior probability, i.e., the probability without considering any contextual factors. This kind of probability is supposed to facilitate the left-to-right processing that occurs both in spoken and visual word recognition. The forward transitional probability was chosen to predict the likelihood of the word combination to occur together since it can help investigate the single-step memory retrieval, i.e., the holistic retrieval of entrenched chunks.

The word combination *erneuerbare Energie* (renewable energy) has the highest transitional probability among the words obtained from the German Web corpus, and the combination *fotographisches Gedächtnis* (photographic memory) has the highest transitional probability based on the student questionnaires. The stimuli for the experiment were chosen based on the students' responses. The criterion for selecting was that at least 50% of the participants had provided the same noun to the given adjective. This resulted in 28 adjective-noun pairs, equally distributed between general and legal word combinations. The words are ranked according to the students' responses. The final list is presented in Table 5.

Table 5. High transitional probability stimuli used in experiment III

General	Students' responses in %	Web TP	Legal	Students' responses in %	BGH TP
fotographisches Gedächtnis	100	0,53	verhängte Strafe	89	0,72
unheilbare Krankheit	100	0,19	versuchter Mord	77,45	0,11
ausgewogene Ernährung	100	0,15	dringender Tatverdacht	70,32	0,47
demograpshischer Wander	97,1	0,46	entlastende Beweise	67,34	0,004

zusammengeknülltes Papier	97,1	0,16	ständige Rechtsprechung	65,53	0,74
schulterlanges Haar	97,1	0,37	verminderte Schuldfähigkeit	63,67	0,36
erneuerbare Energie	94,3	0,83	verdeckter Ermittler	62,75	0,16
sprudelndes Wasser	94,3	0,12	lebenslange Freiheitsstrafe	62	0,61
ohrenbetäubender Lärm	88,6	0,20	begangene Straftat	57,55	0,12
verschlossene Tür	82,9	0,26	vorbestrafte Täter	55	0
missliche Lage	82,9	0,58	angefochtene Urteil	54,75	0,44
weiterführende Schule	71,4	0,23	ermittelnde Staatanwaltschaft	54,75	0,005
alphabetische Reihenfolge	65,5	0,22	zugelassene Revision	52	0,52
tränenreicher Abschied	54,3	0,20	erlassenes Haftbefehl	50	0,05

The bigrams in Table 5 belong to the category of high transitional probability word combinations. To be able to investigate the role chunks and chunking play in the process of entrenchment, two additional sets of bigrams were extracted. The first group is meant to represent low transitional probability word combinations. The nouns from the high transitional probability bigrams were paired with adjectives that were likely to occur with a range of different nouns (e.g., *schlimme Krankheit vs unheilbare Kranheit*) to analyse to what extent transitional probability defines the process of chunk formation. The criterion for selecting the bigrams was that the transitional probability score of the bigram was less than 0,1. The last group of bigrams represented high-frequency word combinations. The adjective-nouns pairs were selected from the German Web and BGH-Urteile corpora. The three groups of stimuli are equally long and include 28 bigrams that fall into one of two groups: law terminology and general words. The rationale behind using three different types of bigrams is to examine the role transitional probability and frequency play in the process of chunking and compare their effect size.

Once the critical bigrams were selected, they were embedded into sentences. First, sentences containing the bigrams were extracted from the German Web corpus. The reason for using corpus-driven sentences was to present participants with sentences that are likely to occur in natural language. However, the sentences were slightly modified to be meaningful and idiomatic without any additional context. A further criterion for the sentence selection was applied: the target bigrams should be in the middle of the sentence to weaken the position effect in the recall task. Two different types of position effects can influence the results of the recall tasks, the primacy and the recency effect. The primacy effect affects the recall of the words at the beginning of lists or sentences. These words and phrases are typically recalled better because they are subject to a greater elaborative rehearsal. In the recency effect, words/phrases from the end of the list or sentence are recalled better because they are still supported by the maintenance rehearsal (Baddeley 1992). Sentence length in all three conditions varied between 15 and 24 words. The reasoning behind not controlling for the length of the sentence was to ensure that sentences were not too short (resulting in that participants could recall all of them without effort) or too long (they could not remember them at all). Additionally, it was ensured that the sentences do not refer to entities that may evoke strong feelings, such as political figures. The word list and the stimuli sentences are available in the appendix.

7.4.4 Experiment IV

The stimuli for experiment IV were very similar to that of experiment III; it consisted of modifier noun combinations elicited from the two corpora. The only difference between the stimuli was that it was divided only into high-frequency and high-probability combinations. The intermediate group with low-transitional probability was removed after piloting. Each of the groups thus was completed with new modifier noun combinations. In the end, the list consisted of two groups of 50 items each, divided between law terminology and expressions used in everyday life. They were selected using the same CQL pattern mentioned in the description of experiment III. Both corpora provide the possibility to rank the query results according to different statistical metrics. The query results were ranked according to log-likelihood to ensure that the newly added stimuli were comparable with those cross-validated using the online student survey. This metric is the closest to transitional probability. After selecting them, the transitional probabilities were calculated for each stimulus, and the frequency of occurrences was extracted from the respective corpora. The updated word combinations and sentences are in the appendix.

7.5 Experimental design

This section will describe in detail each task that was part of the four experiments. The experiment or experiments in which the task was used will be indicated. The main experiments will be described in chronological order before moving on to supplementary experiments.

7.5.1 Read aloud task (experiment I)

The design of this task was rather simple. The selected words described in the stimuli section of the experiment were embedded into a text that resembled a newspaper article about a criminal case. The text was printed on white paper in a dyslexia-friendly font (Arial). The font size was 14, and the line spacing was 2 to ensure that the text and the letters themself appeared less crowded. Participants were instructed to read the text aloud at a natural pace. The participants' voice was recorded using a TONOR TC-777 microphone. The objective of the experiment was to elicit speech similar to what occurs in non-experimental contexts that included the target words. Interviewing the participants would arguably result in a more natural speech than reading a text, but the chance that the participants' answers would not contain the target words is rather high. Reading is an activity that people do on a regular basis, so one can assume that participants will demonstrate temporal or phonetic reduction when encountering an entrenched word and pronounce the words more accurately and slowly if they are not familiar with the word or have not been frequently exposed to it. While reduction was thought to be a feature of only spontaneous speech, more recent experimental research proved that it is also present in speech elicited through reading (Jande 2003).

7.5.2 Rapid word naming task (experiment II)

The design of this particular task was inspired by the claim of Schmid (2017b: 449) that "… access to frequent lexical items in auditory and visual comprehension is among the best candidates for fast, uncontrolled, uncontrollable, and efficient processing". The assumption could be interpreted in the following way: the processing of high-frequency items is automatic and does not reach consciousness. Following this assumption, one could surmise that an experiment where participants have to name words in isolation as quickly as possible could reveal information about the degree of entrenchment of that particular word. This kind of method is usually referred to as a rapid automatic naming task[73], and it is frequently used to diagnose dyslexia. There is a direct correlation between the speed of word/picture naming and difficulties in reading. Theories of lexical access and human memory attempt to describe the underlying processes that happen when a language user is exposed to auditory or written stimuli. The speed of naming can be connected to the ease of access and decoding of words. Schmid (2017b: 450) suggests that articulatory phonetics in spoken production is a perfect candidate for measuring automatic language processing. He defines this process as the routinisation of articulatory gestures, which describes a similar phenomenon that Bybee (2002b) defined as neuromotor automaticity.

The principle of a rapid naming task is that participants are exposed to written words appearing on a screen one at a time and have to read the words as quickly as possible. It is accepted that humans vary greatly in terms of speech production, including the complexity of the sentences they produce and the speech rate with which they utter the sentence. These capabilities are influenced by several social, physical and psychological factors, including the age, gender, and social background of the speaker, as well as education, place of residence and even the psychological characteristics of one's personality. Speech rate is inherently speaker-specific, "pertaining to the inherent speed of articulatory movements which define unique speaker characteristics along with other variables such as voice, use of prosody, or pausing" (Jacewicz et al. 2010: 233). Speech rate varies among speakers as well. This variation can be explained by the length of the utterance, discourse complexity, formality, affect, mood, and communication style in noisy environments or over a longer distance, to name but a few (Jacewicz et al. 2010). To partially eliminate the inter and intra-speaker variation, the participants were asked to pronounce the words as quickly as they could and were reminded of it during the breaks.

The task was designed using the software Psychopy (Peirce et al. 2019). The 60 nouns (30 from the field of law, 30 belonging to general language use) appeared on the middle of a dark-grey screen printed in Arial at font size 48. The list of nouns was

[73] For a detailed overview of the rapid word naming task see Bridge Denckla & Cutting (1999)

randomised for each participant. Each of the words was repeated three times during the experiment. The rationale behind the three repetitions was to measure the development of an articulatory routine and, at the same time, investigate the influence of discourse mentioning.

7.5.3 Sentence reading and verbatim recall task (experiment III + IV)

This particular task was designed to evaluate the effects of different types of long-term linguistic knowledge on verbatim recall and evaluate the linguistic predictors' effect on temporal and phonetic reduction. Sentence recall is usually referred to as sentence repetition and is mainly used to investigate specific language impairments[74] in children. Findings from previous research demonstrate that verbatim recall draws on and benefits from participants' long-term linguistic knowledge (Polišenská et al. 2015). According to Divjak (2019: 131), entrenchment is the process in which linguistic experiences are mentally encoded and committed to memory. Blumenthal-Dramé (2012: 10) suggests that our knowledge of language consists of nodes of entrenched chunks at varying degrees of schematicity and complexity. Chunking can occur by two different means: either through strategic reorganisation based on familiarity or prior knowledge or through grouping based on perceptual characteristics (Gilchrist 2015). Therefore, if entrenchment is a process during which we encode linguistic information into memory, and if our linguistic knowledge consists of nodes of entrenched chunks and verbatim recall measures this long-term linguistic knowledge, this experiment seems to be appropriate to answer the research questions posed in the previous section by experiment III and IV. This method utilises free recall of verbal materials and assumes that the associative strength of items within the same chunk is stronger than between items that are not entrenched. Given that items that are part of the same chunk are more likely to be bound together than items that come from separate chunks, it follows that information that is retrieved in the form it was presented without any kind of modification (only the noun or the adjective is recalled, or they are substituted by another element) must be part of the same chunk.

The materials for these tasks were the sentences containing modifier-noun bigrams described in the section stimuli for experiments III and IV. The experiment used a Latin square design to ensure that the effect of sentence order was eliminated. The task

[74] It is an older term for developmental language disorder

included 84 sentences that were divided into three blocks in experiment III and 100 sentences divided into four blocks in experiment IV. Each block contained an equal number of sentences with different types of bigrams. The high and low transitional probability bigrams had the same nouns. If the high transitional bigram occurred in block A, the low-transitional bigram was only presented in block B or C. This issue was not present in experiment IV because the low-transitional probability bigrams were removed from the stimuli. Additionally, the order of presenting the different types of bigrams was controlled for; half of the stimuli contained first the high-transitional probability bigrams and the other half the low-transitional ones. The experiment has four repetitive parts. As the first step, participants were presented with a sentence that they were asked to read aloud. The sentence was followed by a list of adjectives that were all possible collocations of the noun or synonyms of the adjective in the bigram. After being exposed to the adjectives, the participants were shown another adjective and were asked to decide whether the adjective was part of the list or not. Half of the adjectives were part of the list, and half of them were newly presented. Finally, the participants were asked to repeat the sentence they had read. The adjectives between the reading and recall served to investigate the strength of memory representation for the selected bigrams. At the end of the list, the decision task was used to distract the participants' attention from the fact that adjectives were used to challenge their memory and make the recall cognitively more demanding[75]. The adjectives appeared on the screen one by one and were shown for 1000 milliseconds. The hypothesis that accompanied the design was that the adjectives in the high-transitional probability condition are less likely to be replaced by any of the adjectives from the list than the low transitional adjectives. The list of adjectives that followed the

[75] One of the significant components of Baddeley's multi-component model of working memory Baddeley (2007) is the phonological loop. It is defined by Gathercole & Baddeley (1993: 8) as "a slave system specialised for storage of verbal material". The phonological store represents the material in the phonological code which decays with time. The articulatory rehearsal serves to refresh this decaying information and helps to keep it the working memory. The rehearsal additionally helps to recode nonphonological input such as printed words into their phonological form so that they can be held in the phonological store. In contrast, spoken speech information is immediately directed to the phonological store. A few subjects who were not required to read out the adjectives were rehearsing the sentences, and the information was continuously refreshed in their phonological store. It did not require much cognitive effort to recall the sentence correctly; they only had to execute the phonological plan.

As soon as this tendency became obvious, the instruction of the sentence reading, and recall task was modified to interrupt the process of rehearsal. Subjects were asked to read out the adjectives. As soon as the new phonological forms entered their phonological store, they were not able to rehearse the sentences, and the recall of the sentences became less accurate, included more substitutions and showed a clear preference for recalling the entrenched chunks.

sentences, including the low-transitional bigrams, always contained the high-transitional adjective. The design of the experiment is visualised in Figure 7.

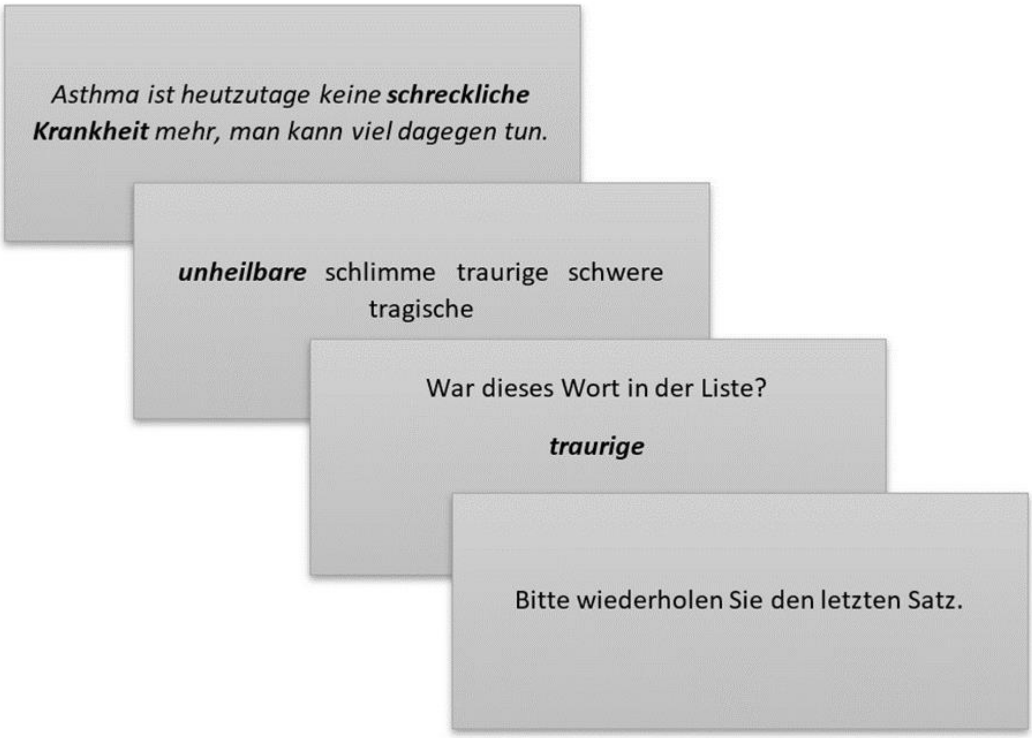


Figure 7. The design of the sentence reading and verbatim recall task

After each block, participants were given a chance to rest or drink some water. The length of the pause between the blocks was determined by each participant according to their individual needs, varying between 2 and 5 minutes. The order of the blocks was mixed in order to weaken the effect of fatigue. The experiment was designed using PsychoPy; the sentences were printed in black in Arial at a font size of 20 points on a light grey background.

Throughout the experiment, the participants' voices were recorded with a TONOR TC 777 digital recorder. The recordings of the sentence reading were used to measure the fluency of speech, which, in this dissertation, is defined as a manifestation of automaticity, the end product of entrenchment (Bybee 2007b, 2002b; Schmid 2017b). The recordings of the sentence recall served to measure the recall accuracy and thus allowed us to investigate the strength of the representation of the chunks.

7.5.4 Self-paced-reading task (experiment II + III + IV)

The experimental battery includes two self-paced reading tasks for experiments II and III, one for experiment IV. The aim of the tasks is to reveal information about the reading habits of the participants. The results of the first pilot study indicated that people find it difficult to accurately report the amount of time they read on average in a week. The reading habits of participants provide critical information for the study because of the logical assumption that increased engagement in reading activities results in a higher chance of being exposed to the target words. Consequently, those participants who demonstrate a higher reading speed should be able to recognise the words more quickly because the time spent decoding a word is determined by exposure to this word. Self-paced reading (SPR) is a computerised method for measuring the reading time for each designated segment of a sentence or series of sentences that are presented as an experimental stimulus (Jegerski & VanPatten 2013). Just et al. (1982) developed three different methods to represent the stimulus on the screen: cumulative condition, moving-window, and stationary-window (for a detailed description of the three methods, see (Jiang 2012). Besides the presentation mode, the number of units shown on the screen simultaneously can influence the experiment's outcome. Besides the word-by-word option, there is a phrase-by-phrase and a mixed option. The benefit of the word-by-word representation is that reaction time is captured for each word separately. However, it is not usual for the participant to read only one word at a time. The phrase-by-phrase method, in contrast, simulates a more natural setting, but the reaction time is captured for the whole phrase. In this particular study, the phrase-by-phrase method was chosen with the stationary-window option for experiment II and sentence-by-sentence for experiments III and IV.

The task aims to measure the participant's reading habits and to determine how familiar they are with the field of law. Two different texts were used for the self-paced reading, a law text and a newspaper article in experiment II and four short stories for experiments III and IV. The law text was taken from Prof. Dr Wolfgang Heinz (2004) publication. An extract of 236 words was used for the self-paced reading task from the above-mentioned article in experiment II. The second text was taken from *Der Spiegel Online*, which is the online version of a German weekly newspaper. The article was published on 15.03.2019 and reports on the Uno Conference in Nairobi. The topic of the article, climate change, should be familiar to most participants as it is a current and relevant topic in the German media nowadays. An extract of 240 words was taken from the article. Five short texts were used for experiments III and IV, each approximately 90 words long. The texts were extracted from the website of *Deutsche Welle*. The texts spanned a range of topics, including the origin of phrasal verbs, the meaning of colours or everyday problems at school. The previous experiment showed that participants outside the law field found it challenging to read the law text and lost interest towards the end of the experiment. In an attempt to make the law text more comprehensible, a *Wikipedia* article on criminal law was used for the experiment. The law text was completely omitted in experiment IV due to the very low comprehension rate of the texts by the non-law participants.

After each paragraph, comprehension questions were asked to ensure that the participants paid attention to the text. Given their general knowledge of the field, the law participants could arguably answer the questions without reading the text. To eliminate this possibility, the comprehension questions were not focusing on the content of the text, as it is typically used in SPR tasks, but rather asked whether a specific word was mentioned in the previous excerpt of the text. This kind of question still requires the reader's attention but can be answered even if someone is not an expert in the field of law. The comprehension questions in the general texts were content-based. The self-paced reading tasks were designed using PsychoPy.

7.5.5 Familiarity judgement tasks (experiments II, III, IV)

Different frequency measures are a vital component of many lexical access models. The effect of frequency is not only widely discussed but is also empirically proven. Frequency information is almost exclusively obtained from large corpora. In this context, the question is whether frequency obtained from the corpora is an adequate index or proxy of frequency in the mental lexicon. It is widely accepted that frequently occurring words in a language are known to most speakers of that language and that they will have a strong mental representation of the words with higher frequency. The low-frequency words are the ones that might raise some concerns. If a word is infrequent in the corpus, it does not mean that all language users will treat the word as a low-frequency word. It has been suggested that word familiarity, a subjective measure based on familiarity ratings, is a more accurate measure. In a series of experiments, Gernsbacher (1984) demonstrated that word familiarity accounted for inconsistent interactions between word frequency and several other lexical variables (bigram frequency, concreteness, and polysemy), particularly for low-frequency words. The familiarity judgment task was included in the test battery for these reasons.

The design of the experiment is simple. In experiment I, the familiarity judgement test asked the participants to judge their exposure to the word. Participants could choose from the following four categories: *I have never heard the word, I have already heard it a couple of times, I am familiar with the word, I actively use the word*. This familiarity judgement test was a paper and pencil task.

In experiments II, III, and IV, participants had to judge on a scale from 1 to 7 how frequently they use the word that appears on the screen. 1 means that they have never heard the word; consequently, they have never used it. 7 means they are familiar with and use the word regularly. In experiment II, the participants were asked to judge both legal and general words. In experiments III and IV, participants gave familiarity scores to the adjective-noun bigrams. There was no time limit on how quickly the participants should carry out the task. The words appeared on the screen one at a time. The words were printed in white on a grey background in Arial at 48 points. The sequence of the words was randomised for each participant. The experiment was implemented in PsychoPy.

7.5.6 Memory tasks (experiment III)

The rationale behind including tasks that measure the participants' working memory is to investigate the correlation between the accuracy of the participants' verbatim recall and memory capacities. Additionally, it serves as a base to explore the role individual differences play in the process of entrenchment.

As mentioned in the self-paced reading task description, Daneman & Carpenter's (1980) test battery is the standard method for testing working memory. Besides the reading comprehension task, their test battery includes reading and word span tasks[76]. The reading span task asks the subjects to read a series of sentences at their own pace and recall each sentence's last word. The number of sentences steadily increased until the participants could not recall all the last words of all the sentences. As this task requires participants to read aloud sentences, and the current study already included a task with the same instruction, this task was replaced by the reading span task developed by Stone and Towse (2015). This task will be referred to as number span to avoid confusion.

[76] Bunting et al. (2006) provide a detailed analysis of the strategies that participants use during memory tasks

7.5.6.1 Number span task

The experiment is designed using the Java-based experimental tool, *Tatool.* [77] The experiment aims to measure the working memory span of the participants during a cognitively complex task. Participants are initially shown a number, which they have to keep in mind; the number is followed by a sentence that might or might not make sense. The subjects have to decide whether the sentence is plausible or not as quickly as possible. All the sentences are grammatically correct, but half of them are not plausible from a semantic point of view. The sentence is followed by another number and then by another sentence.

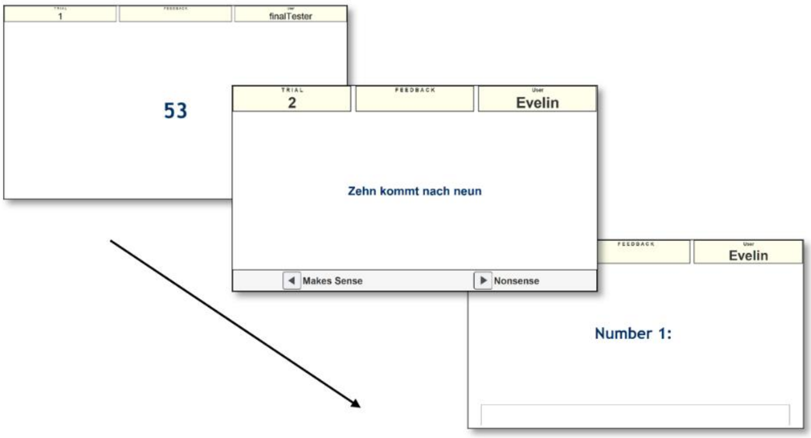


Figure 8. The number span task

After the two sentences, the participants are asked to enter the numbers they have seen. The number of digits and sentences grows after every second trial, with a maximum of 7 digits and 7 sentences. The procedure of the experiment is visualised in Figure 8.

The task measures two different abilities at the same time; first of all, how many different things can the participants simultaneously store in their short-term memory, and second, how quickly they can judge the correctness of the given sentences.

The original experiment was designed in English by Stone & Towse (2015). The instructions and the example sentences were translated into German, and then the program was recompiled with the new sentences.

[77] It is freely available at http://www.cognitivetools.uk/cognition/

7.5.6.2 Word span task

The design of the word span task followed the methodology developed by Daneman & Carpenter (1980). Participants were presented with a set of individual words. The words were presented visually to participants. Research has shown that there is no difference between the visual and auditory presentation of stimuli (Daneman & Carpenter 1980). The test was constructed with 56 frequent monosyllabic words that were not related. The words automatically appeared on the screen in a mixed order and disappeared after 100 milliseconds. The participants were instructed to recall the words in the correct order. The words were grouped into two sets, each set including four, five, six, seven and eight words. Participants were warned to expect the number of words in a set to increase throughout the experiment. The experiment was designed using PsychoPy.

7.5.6.3 Mental processing speed

Besides the three memory tasks that are part of the test battery mentioned above, an additional task was included to measure the participants' processing speed. According to Holdnack et al. (2015: 393), processing speed " is the ability to identify, discriminate, integrate, make a decision about information, and to respond to visual and verbal information". Processing speed provides an estimation of how quickly language users can process information and make specific evaluations, which can indicate the automaticity of that process, accessibility to that information, and efficiency of early stages of information processing. In the literature, there is no established way to measure processing speed. The design used for this study was taken from the online platform *Psychology Today* and modified to meet the requirements of the present study. [78] The task consists of word/image pairs and simple mathematical equations or number sequences. If a pair matches, participants are asked to click the "Correct" button (the left arrow key on the keyboard). If the pair does not match, they have to click the "Incorrect" button (the right arrow key on your keyboard). However, participants had to reverse their answers if the word "Opposite" appeared at the top of the screen. The pictures do not include any violent scenes or inappropriate objects. Participants are asked to perform the task as quickly as possible without compromising accuracy. The rationale behind using this task was to measure how quickly participants

[78] The original experiment is available und the following link: https://www.psychologytoday.com/intl/tests/iq/mental-speed-test-version-1

could process information. It is assumed that there is a correlation between the processing speed and fluency of pronunciation. Higher processing speed can indicate that during word recognition, participants need less time to decode the word; thus, the phonological plan can be executed quicker because the impulse reaches the motor cortex earlier.

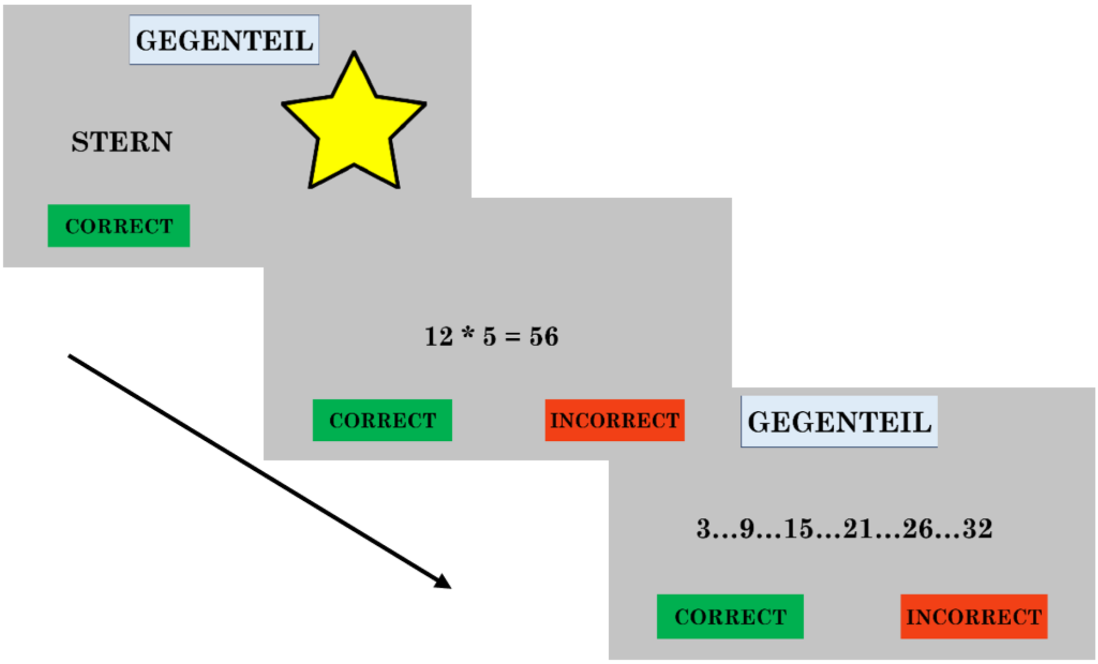


Figure 9. Visualisation of the mental processing task

7.5.7 Implicit and explicit learning skills (experiment IV)

Learning plays a vital role in this dissertation's theoretical framework on entrenchment. Speakers differ in their explicit and implicit learning skills. The moot question is whether these learning skills influence language users' ability to entrench specific linguistic structures. Ellis (2015: 2) provides the following definition of the two types of learning: "implicit learning is the acquisition of knowledge about the underlying structure of a complex stimulus environment by a process which takes place naturally, simply and without conscious operations" and "explicit learning is a more conscious operation where the individual makes and tests hypotheses in a search for structure". The definition of implicit learning underscores that this type of learning happens without attention and consciousness. In other words, if we follow the definition of automaticity, this type of learning happens automatically. Accordingly, explicit learning could be labelled as a controlled process. In a recent study, Llompart & Dąbrowska (2020) found that scores on explicit memory tests significantly correlated with the participant's grammar, vocabulary, and collocation knowledge. In experiment IV, the participants' explicit and implicit learning skills were tested to investigate how these skills influence the process of entrenchment. Explicit learning is measured with a paired-associate learning task, and implicit learning skills are tested using the embedded triplets task.

7.5.7.1 Paired-associate learning task

Paired-associate learning is a classic memory paradigm that measures how people encode and retrieve newly formed associations among stimuli. In a typical study using paired-associate learning, people are asked to learn unrelated word pairs (e.g., *stove – letter*). After the participants have been introduced to all the pairs, memory for those pairs is typically tested by having them recall one of the words in response to the word they were paired with during encoding (e.g., recall the word paired with "stove"). Another way of testing the memories is to ask the participants to distinguish between word pairs that were encoded together (e.g., *stove – letter*) and word pairs composed of two words that were studied but were not paired during encoding (e.g., *stove – dance*; known as associative recognition) (Arndt 2012).

For experiment IV, 60 frequent but unrelated words were selected and randomly paired. The words were extracted from the GermanWeb corpus. Only monomorphemic words were used. During the piloting of the task, participants were presented with 30 pairs, after which the testing phase took place. Due to the poor performance of the participants, the task was later divided into three blocks, and participants were tested for their memory after 10 pairs. Participants were asked to distinguish between pairs they had learnt and unseen pairs. The experiment was designed in PsychoPy. Each word pair was shown for 2000 milliseconds to make the task more challenging.

7.5.7.2 Implicit learning task

There are numerous ways to measure the language user's implicit learning skills. A common practice is to let participants watch a continuous stream of evenly paced, individually presented items on a monitor and, later, surprise the participant with questions about what they saw during this familiarisation phase. Typically, stimulus presentation times are consistent and are set at somewhere between 200 and 800 ms per item. Another possible method is to organise the input randomly in triplets during the familiarisation and, during the testing phase, ask participants to do a forced-choice task to measure whether they have implicitly learnt the triplets (Arciuli & Simpson 2012).

For experiment IV, the latter methodology was used. The design is based on the methodology used by Arciuli & Simpson (2012, 2011)[79]. The stimuli consisted of 4 pairs of triplets and 12 distractors. The triplets and the distractors were visualised as colourful aliens[80]. The experiment started with a familiarisation phase during which participants were exposed to a continuous stream of these aliens. For a single participant, each alien was visible for the same length of time (400 ms). A blank screen was presented for 200ms between each alien. Each base triplet was selected for inclusion in the familiarisation stream 24 times each (giving a total of 96 triplets). For six of these 24 instances, one of the aliens was presented twice in a row to provide a cover task (the Alien Alert game) that would ensure participants paid attention during the Familiarization Phase (participants were required to press a button whenever they saw a repeated alien). The repetitions of the aliens were counterbalanced within each triplet. So, for example, for base triplet ABC, there were two occurrences of AABC, two of ABBC, two of ABCC and 18 of ABC. Although there were two occasions when the strict triplet structure was violated (e.g., ABBC), there were still 22 occasions when the stimuli appeared in strict triplet order. Thus, the familiarisation stream consisted of 312 individual aliens, with each of the 12 appearings 26 times each. The order of the triplets within the familiarisation stream was randomised.

The test phase followed the familiarisation phase. Four new triplets were created, with each containing one alien from three different base triplets. Due to ordering constraints, these new triplets never appeared in the familiarisation stream and are referred to as impossible triplets. One base triplet was displayed with one impossible triplet for each test trial. The aliens in each triplet were presented one at a time using the same presentation time as used in the Familiarization Phase with a 1000 ms gap separating the two triplets. After all six aliens had been presented, participants were prompted to identify which of the two triplets had appeared previously (during

[79] Many thanks for Ashley Blake from the University of Birmingham for sharing her embedded triplets experiment with me.

[80] Ashley Blake designed the experiment to test the implicit learning skills of children which was the reason for using colourful aliens.

familiarisation). To make the task easier, one of the groups of triplets appeared next to a spaceship and the other group next to a rocket. Participants were asked to indicate which triplet they had already seen. This constituted a forced-choice task. No time constraints were imposed, and responses were made by clicking either on the rocket or the spaceship. The presentation order of base triplets and impossible triplets was counterbalanced. Across the 64 test trials, each base triplet and each impossible triplet were seen 16 times. Thus, if any additional implicit learning took place during the Test Phase, opportunities to learn were equal for both types of triplets. Each participant received a different random order for the test trials. The experiment was designed in the online experimental platform Gorilla.

7.5.7.3 Cued task switching (experiment IV)

The task-switching paradigm provides a laboratory tool for studying mental flexibility and goal-directed behaviour control in situations involving several tasks (Monsell 2003). This paradigm requires participants to switch frequently between two or more tasks. Switching from one task to another is slower and more error-prone than repeating the same task, and this persists even over the massive practice (Vandierendonck et al. 2010). Seemingly, the act of switching requires special mental operations, and the task-switching paradigm is designed to investigate the nature of these operations. The present study used task switching to measure processing speed (the participants' reaction time) and executive attention (accuracy). Participants were introduced to two different tasks: a shape task and a colour task. The stimuli consisted of a circle and a rectangle in the shape task. If the participants saw a circle, they were instructed to press the letter *b* on the keyboard, and in the case of a rectangle, the letter *n*. In the colour task, if the stimulus was yellow, they had to press *b*, and if it was blue, then *n*. In each trial, participants were first shown a fixation cross, which was followed by either the word "shape" or "colour". The word indicated whether they had to react to the stimuli's shape or colour. The cue (i.e., "shape" or "colour") was displayed for 350 ms. The interval between the cue and the stimuli was set to 1000 ms. The familiarisation phase included 5 repetitions of each condition. The test phases consisted of 50 trials, equally subdivided between the "shape" and "colour" tasks. The experiment was taken from the experimental library of Psytoolkit[81] (Stoet 2010; 2017)

---

[81] The experiment is free available under the following link:
https://www.psytoolkit.org/experiment-library/taskswitching_cued.html

7.6 Participants

In the four experiments, 126 participants were tested. These participants fall into two groups, participants related to the field of law, and the other half were recruited to have as little connection to the area of law as possible. In experiments I, III, and IV, mostly younger subjects were tested who either studied law and attended the professors' lectures whose materials were used to select the target legal words or students in fields unrelated to law. In experiment II, employees of two law chairs were recruited, including professors with over 30 years of experience, post-docs, graduate students, and student-helpers. To make the groups as homogenous as possible, the non-law group also included professors, post-docs, and graduate students from other fields of study. Data about the participants are summarised in Table 6.

Law participants were recruited in the library during the semester break and through social media. The non-law group was gathered mainly through personal contacts. None of the participants was familiar with the tasks or the research's aim. Eligibility criteria required individuals to be native speakers of German and be enrolled as a student in the law department for the law group, and have no connection to the field of law for the non-law group. All the participants received compensation for their participation. In experiment II, the participants were asked to indicate how long they had been studying law or working in this sphere. The law group's mean exposure was 10.78 years; the youngest participant had only 3 years of exposure, and the oldest 30 years. In experiment III, two participants in the initial stage were tested to examine the proper instruction for the sentence reading and recall task; therefore, their data will not be analysed. This resulted in 14 participants. In experiment IV, participants were selected to be between the ages of 20 and 30 because the experiment included a sentence recall task, and some researchers claim that memory capacity starts decaying, albeit slowly, after the age of 30.

Table 6. Summary of age and gender of participants

	Experiment I		Experiment II		Experiment III		Experiment IV	
	Law	Non-law	Law	Non-law	Law	Non-law	Law	Non-law
Number	8	8	18	18	7	9	23	39
Age	19-26	26-49	20-48	21-47	18-31	20-34	19-28	20-31
mean	22.5	39.3	29.5	29.6	23.7	25.3	23.13	23.38
Gender	7 females		17 females		10 females		41 females	

7.7 Procedure

Friedrich-Alexander-University's Ethics Committee approved the study. Participants were informed before the experiment that their voices were going to be recorded.

7.7.1 Experiment I

The participants were asked to read aloud a text printed on an A4 sheet in Arial 12 font. The participants were instructed to read the text as naturally as possible. The voice of the participants was recorded with a TONOR TC 777 digital recorder. The recording took place in a silent room. The background questionnaire was completed at the end of the experiment. All the participants gave written consent for their voices to be recorded during the experiment. The participants' consent was obtained before the start of the experiment.

7.7.2 Experiment II

Before the experiment started, the participants were informed about the study and what they were going to do. After that, if they agreed to participate, they signed written consent allowing the experimenter to record their voice. The session always started with the word naming task to weaken the effect of priming.

The word naming task started with an introduction informing the participants that they would see 180 words. The task was to pronounce them as quickly as possible. Before the words appeared on the screen, a fixation cross was shown for 400 milliseconds. A word followed the fixation cross. The participants had to press the space bar to see the next word. The words were presented in random order and were printed in white in Arial, the font size 48, on a grey background. After 60 words, there was always a 10-second-long pause. The words were repeated three times in different orders. This allowed us to measure how the speech rate of the participant changed across the three repetitions. The participants' voices were recorded with an Olympus DS50 digital voice recorder. The participants were sitting in a quiet room, and the experimenter was seated behind them not to distract them but still be able to advise them if needed. The word naming task was followed by self-paced reading. Half of the participants were first confronted with the law text and the other half with the journal article to ensure that the obtained results were not biased because of the order effect. The experiment started with the instructions. The participants were informed that they would be presented with a text segmented into chunks. Their task was to read the text as quickly as possible. The space bar was used to determine the speed of appearance of the chunks on the screen. After two sentences, a comprehension question was asked. The left and right arrows of the keyboard were used to answer the question. The keys were marked with green and red stickers to prevent confusion about which key corresponds to which answer. Before a new sentence started, a fixation cross was shown in the centre of the screen to indicate where the next chunk would be located. The law text was segmented into six parts and the news text into seven. Respectively there were six and seven comprehension questions. The phrases were printed in white on a grey background in Arial 22 and presented in the monitor's centre. The questions were printed in yellow to attract the participants' attention. The participants were informed to read the text as quickly as possible, but they had as much time as they needed to answer the questions.

The session ended with the familiarity judgement task. The participants were exposed to the target words and the distractors again and were asked to judge how familiar they were with the items and how often they used them. The words appeared in random order, and the participants were not limited in time. The session typically lasted 25 minutes.

7.7.3 Experiment III and IV

The experiment was carried out in a computer room. Participants were informed about the general aim of the experiment, and it was communicated that one of the experiments included a voice recording. If participants agreed to participate, then they were asked to sign the written consent. Participation in the experiment was anonymous. Each participant received a personal code that was used to match their responses in the different tasks of the experiments. Each session started with the sentence reading and recall task. During the sentence reading task, participants were instructed to read in their natural manner and not to spend time trying to memorise the sentence; they were reminded of this during the pause between the blocks. Additionally, it was emphasised that the sentences vary in length and complexity, and they may remember some of the sentences more accurately than others. Initially, participants were asked only to read the sentence out loud. After each session, participants were asked to comment on the experiment and give some information about their strategies to remember the sentences accurately. A clear tendency was observed where participants were trying to constantly rehearse the read sentences and only visually perceive the list of adjectives without accessing their meaning or subvocalising them. This strategy led to a ceiling effect because participants were rehearsing the sentences. The solution to how to avoid this constant rehearsal was found in the work of Baddeley on working memory (2007).

One of the significant components of Baddeley's multicomponent model of working memory (2007) is the phonological loop. It is defined by Gathercole & Baddeley (1993: 8) as "a slave system specialised for storage of verbal material". The phonological store represents the material in the phonological code which decays with time. The articulatory rehearsal serves to refresh this decaying information and helps to keep it in the working memory. The rehearsal also helps to recode nonphonological input, such as printed words, into their phonological form to be held in the phonological store. In contrast, spoken speech information is immediately directed to the phonological store.

Subjects who were not required to read the adjectives out loud rehearsed the sentences, and the information was continuously refreshed in their phonological store. Recalling the sentence correctly did not require much cognitive effort; they only had to execute the phonological plan.

As soon as this tendency became apparent, the instruction of the sentence reading and recall task was modified to interrupt the rehearsal process. Subjects were asked to read the adjectives out loud. As soon as the new phonological forms entered their phonological store, they were not able to rehearse the sentences, and the recall of the sentences became less accurate, included more substitutions, and showed a clear preference for recalling the entrenched chunks.

The other experiments were presented to the subjects in a mixed order to weaken the effect of fatigue. Before each experiment, participants were orally informed about the task and were told that they were free to stop the experiment any time they felt uncomfortable. During the reading experiments, participants' attention was drawn to the fact that they should read the texts as quickly as possible but not at the expense of comprehension. Each experiment ended with a background questionnaire to elicit information about gender, age, field of study, and reading habits. Depending on the participants' speed, the sessions last between 1,5 and 2 hours.

8 Data analysis

8.1 Temporal reduction

8.1.1 Experiment I

In the interest of readability, predictor variables such as FREQUENCY will be henceforth printed in SMALL CAPS. The recordings made during the read-aloud task were processed with the WebMouse application that is hosted on the website of the Institute of Phonetics and Speech Processing of Ludwig Maximilian University. The application segments an audio file into SAM-PA phonetic segments given an orthographic transcription. The output file contains the timestamps for each word, followed by the SAM-PA transcription and, finally, the length of each sound.

In order to analyse the results of the two groups, the general speech rate and the phonation rate, as well as the speech rate during the pronunciation of the target words, were calculated for each participant. The PRAAT speech rate script was combined with a Python script that calculated the number of syllables in the text. For this particular experiment, the word's underlying syllables were taken into account. The phonation rate was measured with two different pause settings, first with 300 milliseconds and then with 400[82]. The two different pause settings did not affect the phonation rate significantly ($p > 0.5$).

The speech rate during the pronunciation of the target words should be an indicator of the degree of entrenchment and show whether familiarity with the word can reveal something about neuromotor automaticity. However, the speech rates cannot be compared in this form because they heavily depend on the average speech rate of the participants. A new unit of measurement had to be calculated, which enabled the comparison between the groups and the individuals. The SPEEDUP describes how quickly or slowly the participants pronounced the target words in comparison to their general speech rate. This was calculated by dividing the average speech rate by the target word speech rate (e.g., 7.5 syll/sec / 6.06 syll/sec = 1.2376).

The SPEECH RATE, SPEEDUP, and length of the pronunciation (LENGTH) of the words under investigation are summarised in Table 7. below. As shown in Table 7, the difference between the target and control groups regarding the average speech rate and speed up is relatively small. Numbers in the SPEEDUP column that start with 0 indicate a

[82] De Jong & Rutger Bosker (2013) provide a detailed description on choosing thresholds to measure silent pauses.

slower speech rate during the pronunciation of the target words; variables starting with 1 respectively indicate a speedup.

Table 7. Summary of the speedup, average speech rate (syllable per second) and the length of pronunciation in milliseconds

Group		Mean	Standard deviation	Median	Minimum	Maximum
Speedup	law	1.21	0.21	1.19	0.73	2.02
	non-law	1.18	0.21	1.19	0.65	1.64
ASR (syll/sec)	law	6.2	1.13	6.19	3.73	10.93
	non-law	5.44	0.96	5.35	3.07	7.77
PronLength (msc)	law	1073.92	201.34	1070	640	1700
	non-law	1238.54	260.67	1210	790	2278

An independent, one-tailed t-test was performed to see whether the difference between the two groups' performance was significant. Contrary to expectations, the t-test did not show a significant difference between the SPEEDUP of the law and the non-law group ($p > 0.05$). The obtained results could be due to the test battery or the sampling of the participants. Thus, as a unit of measurement of neuromotor automaticity, SPEEDUP did not meet the expectations. Another independent, one-tailed t-test, which compared the length of the pronunciation of the target words among the two groups, showed a statistically significant difference ($p < 0.001$). Following the present results, SPEEDUP did not reveal the expected findings because of the age difference between the two groups or the small sample size. The mean age of the target group is 22.5, and of the control group, 39.7. The statistically significant difference between the average SPEECH RATE between the groups supports the claim mentioned above ($p< 0.001$, law group = 5.14, non = 4.62). All the differences between the three measurements are shown in Figure 10.

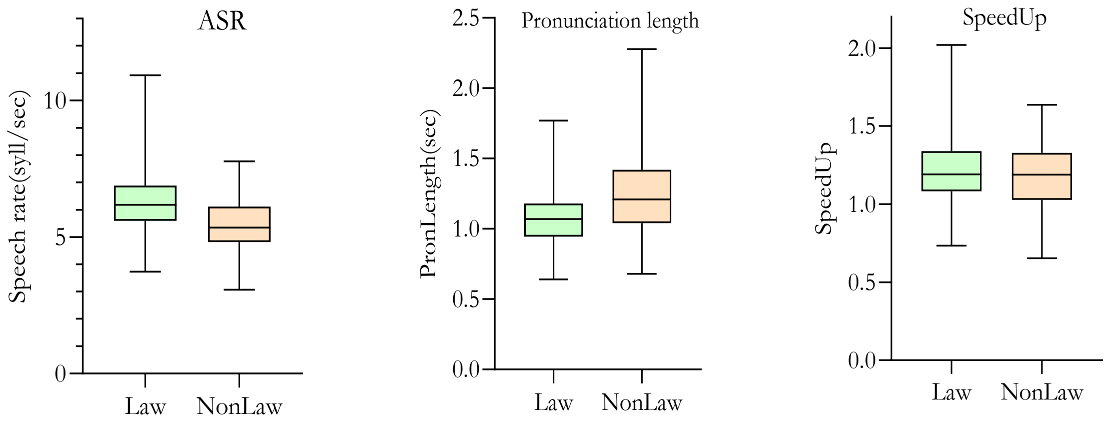


Figure 10. Boxplots of speedup (right), average speech rate in syll/sec (left), length of pronunciation (middle)

Previous studies (Bybee 2007a, 2002a; Pierrehumbert 2002; Jurafsky et al. 2001b) have reported that frequency is one of the main predictors of temporal and phonetic reduction. However, it has to be highlighted that the frequency effect is mainly observed with high-frequency words. The stimuli selected for experiment I are very specific and have a high frequency in the lecture corpora but a rather low frequency in the DWDS.

Figure 11 demonstrates how participants reacted to the different target items. For the law group, the words *Beweisverwertungsverbot, Strafverfolgungsbehörde, Zeugnisverweigerungsrecht* and *Klageerzingungsverfahren* caused the most variation; additionally, these words were pronounced the most slowly. What concerns the non-law group, the word *Zeugnisverweigerungsrecht* has a special status. It was pronounced the slowest and displays the most considerable variation, ranging from 1300 ms to 1650 ms. The word *Beweisververtungsverbot* was pronounced as well slower in comparison to other words. These results could be due to the general low frequency of words in the DWDS corpus. Adjusting the Kucera (Burgess & Livesay 1998) frequency thresholds, all the words used for this pilot study fall into the low-frequency bin. A possible explanation for the slower speech rate, besides the low frequency of the words, could be explained by the lack of familiarity. 6 participants out of 8 indicated that they had heard the words, *Beweisverwertungsverbot* and *Zeugnisverweigerungsrecht* only 1 or 2 times. In comparison, participants from the law group said they had been exposed to the words frequently and had used them quite often, which was expected as all the participants from the law group had attended the lecture from which the words were selected.

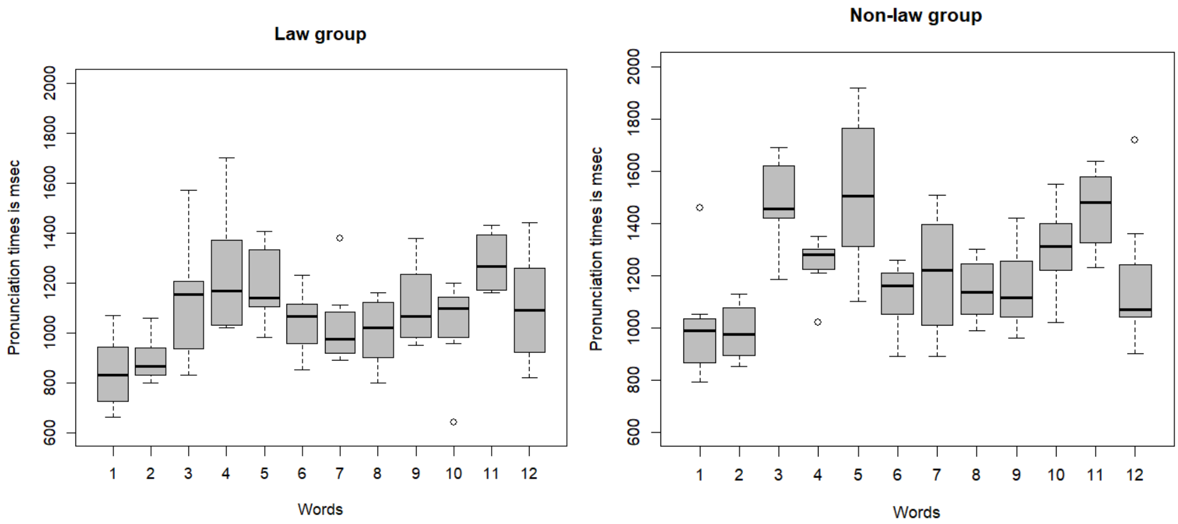


Figure 11. Pronunciation length for the target items. The numbers correspond to the following words: 1- Staatsanwaltschaft, 2 - Ermittlungsverfahren, 3 - Beweisverwertungsverbot, 4 - Strafverfolgungsbehörde, 5 - Zeugnisverweigerungsrecht, 6 - Ermittlungsmaßnahmen, 7 - Verfahrensverzögerung, 8 - Bundesverfassungsgericht, 9 - Rechtsmittelverfahren, 10 - Gerichtsverfassungsgesetz, 11 - Klageerzwingungsverfahren, 12 - Beschuldigtenvernehmung.

8.1.1.1 Modelling temporal reduction

A linear mixed-effects regression analysis was performed to examine how the selected explanatory variables affect the rate of temporal reduction. The explanatory variables were selected by a stepwise selection and the best subset method; their results were compared.

For analysis, the function lmer () from package lme4 was used. First, a baseline model was fitted with intercept only. The model was fitted with the maximum likelihood method. Participants and items were treated as random effects. The explanatory variables were added stepwise, including the LOGFREQNOUN, OCCUPATION, AGE and WORDLENGTH. The obtained models were then compared with the help of ANOVA. The model with the highest explanatory power (R2 =0.309, adjusted R2=0.294) included OCCUPATION, WORDLENGTH, LOGFREQNOUN of the words in DWDS as explanatory variables.

Table 8. Results of the mixed-effects regression model, with pronunciation length as the dependent variable

	Estimate	Std. Error	Df	t-value	p-value
Intercept	0.161463	0.2284	10.2619	0.707	0.49
Occupation:Non	0.164871	0.0518	13.9967	3.178	0.006
WordLength	0.044172	0.0098	9.6858	4.474	0.001
LogFreqNoun	-0.023451	0.0095	9.50266	-2.469	0.03

The results of the mixed-effects linear regression models could be interpreted in the following way: the fact that somebody belongs to the group of lawyers or control group can indicate a 164.87-millisecond difference in the length of pronunciation, which is the difference between the mean length of pronunciation among the two groups. The way to read the output for the WORDLENGTH is that every increase in the number of letters by 1 adds 44.17 milliseconds to the length of pronunciation. Finally, every unit of frequency decreases the pronunciation length by 23.45 milliseconds. Thus, the results indicate that occupation had the most significant impact on pronunciation length. This implies a significant difference between the groups, and SPEEDUP, as a possible measure of automaticity, failed to display these differences due to the small number of observations.

8.1.2 Experiment II

The hypothesis behind the study was that the group of lawyers would pronounce the legal terms more quickly than the control group due to more developed neuromotor routines. Experiment I indicated that there are significant differences between the groups concerning the rate of temporal reduction. The rationale behind this assumption is that familiarity with an item leads to easier access from the mental lexicon and neuromotor automaticity in the Bybee-an sense (2002a, 2002b). However, taking into account only the participants' occupations would not be sufficient to analyse the temporal and phonetic reduction and their underlying processes. Besides OCCUPATION, as a possible predictor for temporal and phonetic reduction, other extralinguistic and linguistic factors were analysed.

The extralinguistic factors include the participants' AGE, GENDER, EXPOSURE TO LAW, and READING habits. As was mentioned above in Chapter 7.5, researchers have established a strong negative correlation between age and speech rate.

AGE of the participants was coded as a continuous variable. AGE ranged between 21 and 48 and was comparable across both groups.

GENDER as a predictor of speech rate is not as extensively researched as AGE. Byrd (1994) claims that men actually speak somewhat faster than females. Yuan, Jiahong, et al. (2006) have obtained small but significant differences between the speech rate of males and females. The gender distribution among the participants is almost equal, 15 male participants and 17 female participants. This allows one to take gender into account as a possible factor that can affect temporal phonological reduction.

EXPOSURE to law as a factor measures how many years people have been exposed to law. It includes their education and work-related experience. This factor, though, contains information only from the law group as all the members of the non-law group were recruited not to be related to the field of law. Nevertheless, this predictor is essential to analyse whether the time one was exposed to the words will have a continuously growing effect on the entrenchment of the word in the mental lexicon and on the development of neuromotor automaticity. Therefore, people with more experience in the field of law should be able to show a more significant temporal reduction. READING habits of language users have not been studied in connection to reduction processes so far. The rationale behind analysing this factor is that people encounter words not only through the auditory channel but also through the visual one. The more time people spend reading, the higher the chance that the target words are entrenched in their mental lexicon due to the frequency of exposure. Information on the reading habits of the participants was obtained with the help of the self-paced reading task. The results were coded as continuous variables representing the time the participants had read the selected texts. Additionally, the type of text was coded as a categorical value (legal text, general text). Figure 12 below illustrates the reading habits of the participants.

Figure 12 is quite revealing in several ways. First, as expected, the lawyers read the legal text significantly faster than the control group members. Second, the non-law group was marginally but not significantly ($p > 0.05$) faster in reading the general text than the law group. Third, the non-law group was 1.4 times faster reading the general text than the law text. The last one indicates that members of the control group are rarely exposed to texts containing legal terms, which therefore indicates a weaker memory representation.

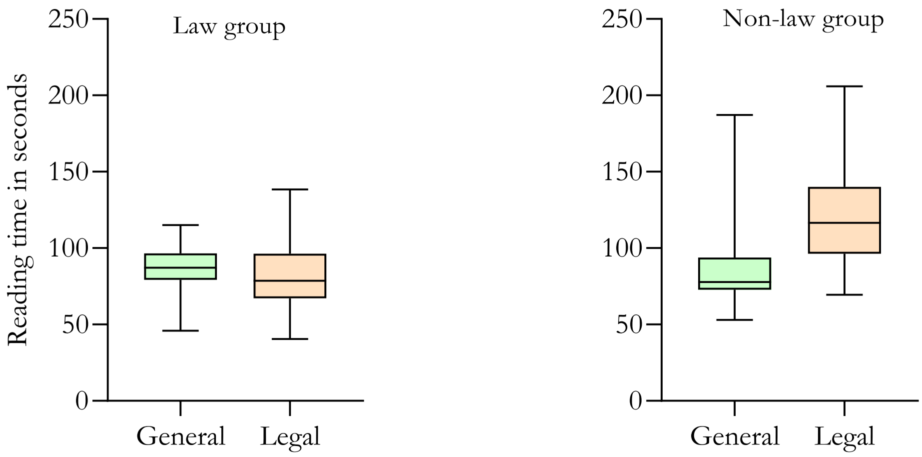


Figure 12. Reading habits of participants in experiment II

Supplementary to the extralinguistic factors, a range of variables were analysed that focused. Among them: the ITEM TYPE, FAMILIARITY, FREQUENCY, WORDLENGTH, the number of times that word has been pronounced (further REPETITION), DISPERSION and the TRANSITIONAL PROBABILITY (TP) on the morphological level[83] of the compound, the relative frequency of the base (LOGRELFREQBASE) and the root of the compounds. The FREQUENCY of the words was elicited from the DWDS, German Web[84] (FREQDWDS, FREQWEB), and the lecture-based corpora[85]. It was coded as a continuous variable, and later it was transformed into log frequencies. These frequencies refer to the frequency of the compounds. Additionally, the frequency of the base of the compound was extracted from the GermanWeb corpus (FREQBASE) and the frequency of ending (FREQSTEMENDING). To analyse the correlation between WORDLENGTH and FREQUENCY, the selected words were sorted into length bins of *short, medium, long, very long*. The bins contained an equal number of words and always increased by three letters, following the study of New et al. (2006).

Moreover, the frequency elicited from the GermanWeb corpus was sorted into frequency bins. The size of the frequency bins was determined using the Francis and Kucera method of establishing the frequency thresholds and adjusted to the size of

83 Vast majority of the words used as stimuli were compounds. By transitional probability the likelihood of the free morphemes that make up the compound occurring together is meant.

84 The German WebCorpus (deTenTen) is a corpus made up of texts collected from the internet. The corpus belongs to the TenTen corpus family which is a set of the web corpora built using the same method with a target size 10+ billion words. The corpus is hosted on the website of SketchEngine.

85 Because it includes only the legal words, this explanatory variable could be added to the regression models.

the GermanWeb corpus (Burgess & Livesay 1998). The results are shown in Figure 13 below.

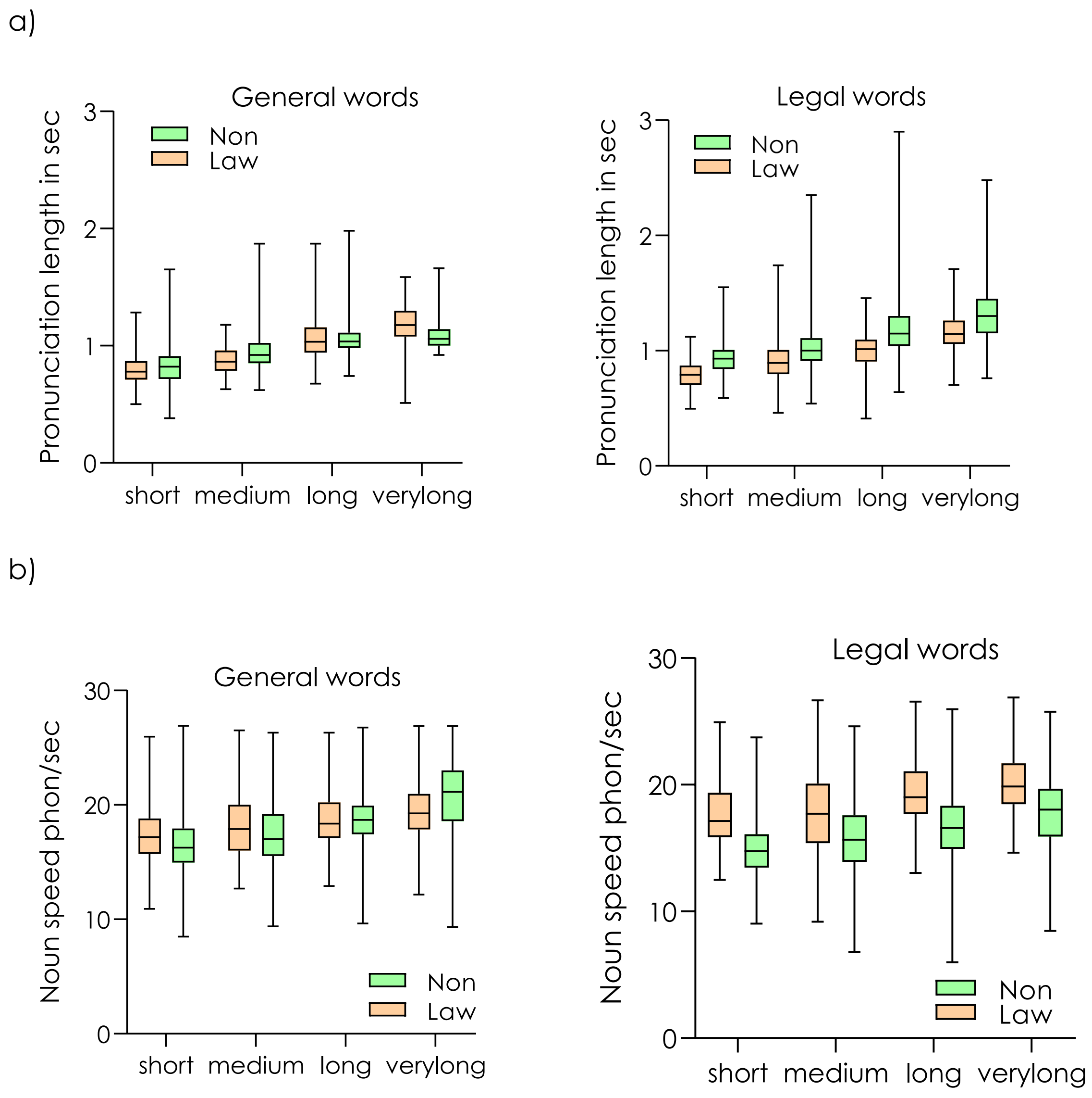


Figure 13. Correlation between pronunciation length and word length (upper row), the correlation between speech rate and word length (lower row). The beige colour represents the law group, and the green the non-law

Figure 13 provides information about the expected correlation between WORDLENGTH and PRONLENGTH. What is more interesting is the two lower figures. The upper figures reveal the longer words were pronounced longer. However, when the length of pronunciation is converted into speech rate, the longest words had the highest speech rate. In longer words, more vowels can be reduced, which leads to a faster speech rate. Research (Baddeley & Hitch 1974) has also demonstrated that the word length effect is in negative correlation with memory recall; shorter words are recalled more accurately than longer words. An additional factor that is likely to influence the words' speech rate and pronunciation length must be mentioned. The law words are sorted into the low and low-medium frequency bins following the frequency threshold of Kucera and Francis (Brysbaert & New 2009). The frequency effect is observed among the high-frequency words in the vast majority of cases. This low frequency might be later reflected in the mixed-effects model results. The words were controlled for frequency in the lecture-based corpus during the data collection phase. This was justified because it aimed to recruit participants from the two law chairs from where the lectures originate. It was assumed that the frequency of the words in the lectures-based corpus would reflect the mental representation of the members of the chair more accurately than in a general corpus. Furthermore, the general words were selected by length to match the terminology, which restricted the possibility of controlling their frequency. The lecture-based frequency of the words was not included in the two global models since it only includes legal terms.

In addition to the frequency of the words, information was gathered on how familiar each participant was with the target words and distractors. FAMILIARITY scores were converted to Z-scores to make comparisons of relative ratings possible. This transformation is relatively common in acceptability judgments (Verhagen 2020), as it involves no loss of information on the ranking nor at the interval level. A score of 0 indicates that a participant judges a particular item to be of average familiarity compared to the other items. A positive Z-score indicates that the raw score is higher than the mean average, and, respectively, a negative score is lower than the mean average.

ITEM TYPE indicates whether the word belongs to the category of legal terminology or general words.

Almost all the words that occur in the experiment are compounds. Some of the words share the same root, e.g. *Verfaren* in words *Verfahrensrüge*, *Verfahrenshindernis*, or *Ermittlung* in words *Ermittlungsrichter*, *Ermittlungsmaßnahmen*, *Ermittlungsverfahren*. This fact led to three different measurements that aim to explore the status of compounds in the mental lexicon. Transitional probability is mainly used to measure the likelihood of one word being followed or preceded by another one. Jurafsky et al. (2001a) investigated the effect transitional probability has on phonetic reduction. The transitional probability of a particular target word w_i given a previous word w_{i-1} is estimated from a sufficiently large corpus by counting the number of times the two words occur together $C(w_{i-1}\, w_i)$ and dividing by $C(w_{i-1})$ the number of times that the first word occurs. The bigram frequency was replaced by the frequency of the compound and divided by the frequency of the stem with all other possible endings. The query to determine the frequency of the stem with all possible endings is defined as [lemma="Ermittlung.+"]. The frequency information was extracted from the GermanWeb corpus, and this explanatory variable is referred to as TP-COMPOUND.

The last frequency related measure, i.e. RELATIVE FREQUENCY was adopted from the work of Blumenthal-Dramé (2012). Her study has shown that the RELATIVE FREQUENCY was the only corpus-drive frequency measure that predicted the degree of entrenchment. RELATIVE FREQUENCY is a composite measure that results from dividing the surface frequency of the compound by the base frequency of the word (2012: 112), which, in this case, gave the same results as the transitional probability of the compound. In the analysis, only TRANSITIONAL PROBABILITY was used to avoid multicollinearity.

The last factor used in the analysis is the number of times each participant had to pronounce the word, henceforth REPETITION. It was fixed for all the words and all the participants. The prediction was that there would be a steady decrease in the length of pronunciation with every repetition. The effect of repetition was the same for both groups with varying degrees. Figure 14 below represents this effect.

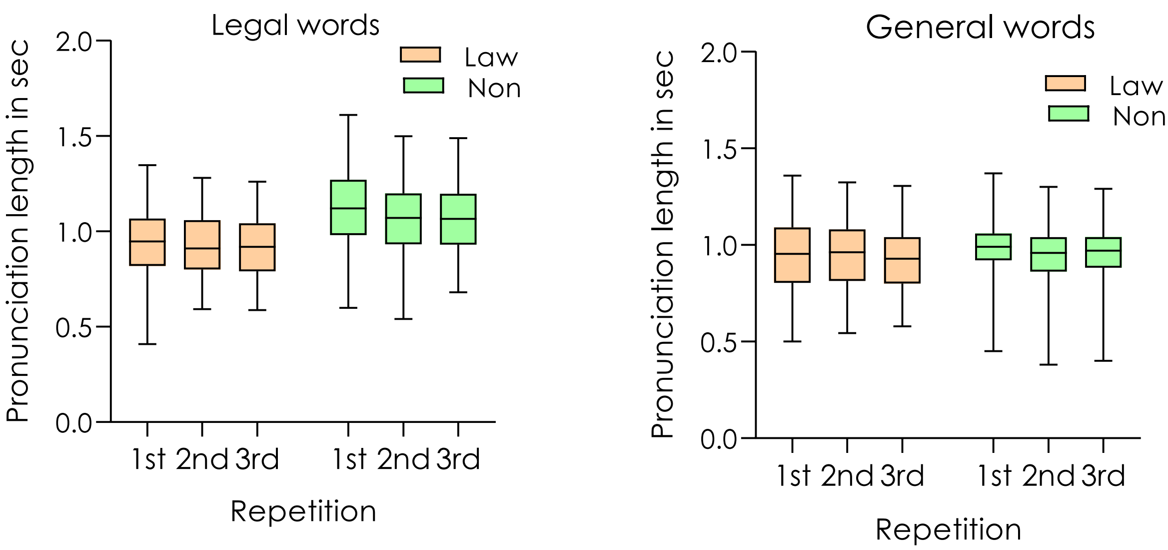


Figure 14. Pronunciation length across the three repetitions. The left-hand-side figure represents the law words lawyers, and the right-hand-side figure shows the general word. Beige colour represents the lawyers and the green the non-lawyers

Figure 14 invites the following conclusions. First of all, the lawyers pronounced the legal words slightly faster than the general words (legal words = 1.018, general words = 1.036, p = 0.08), though the length of the words was controlled across the group of words. Across the trials, the lawyers showed an increased temporal reduction of both types of words. However, the law words were always pronounced slightly more quickly than the distractors. The most significant difference in the length of pronunciation is between the first and the second repetition. This means that people are not capable of exceeding a specific speech tempo. As expected, the members of the control group pronounced the legal words significantly more slowly than the general words (legal words = 1.225, general words = 1.154, $p < 0.001$). This distinction in the length of pronunciation is observable across all three trials. The gap between the pronunciation of the legal terms and the general words is more significant than in the case of the lawyers. Participants' performance improved significantly from the first to the second repetition, especially regarding legal words (repetition I = 1.225, repetition II = 1.151, $p < 0.001$). The pronunciation time continued to decrease towards the third repetition but not as significantly as between the first and the second (repetition II = 1.151, repetition III = 1.371, p = 0.38). Both groups were significantly quicker during the second repetition than during the first. However, the independent t-test shows that the members of the non-law group show a more significant temporal reduction than the law group. The law group, on average, was 41 milliseconds quicker in the second repetition, and the non-law group 77 milliseconds faster. The difference between the two groups can be accounted for by the effect of recency on low-frequency words.

8.1.2.1 Modelling of temporal reduction

The first step in the analysis of temporal reduction was to determine whether the difference between the two groups was statistically significant. The two groups' pronunciation length was compared with an independent, one-tailed t-test.

The boxplot below indicates that there is a significant difference between the two groups. The independent t-test confirmed the hypothesis ($t = -24.049$, $p < 0.001$, df = 5324.5) and demonstrated that there is indeed a difference between the performance of the two groups. In the experiment, participants differed in their speech style, though every participant was instructed to name the words as quickly as possible. Figure 15 illustrates the difference in the length of pronunciation between the two groups.

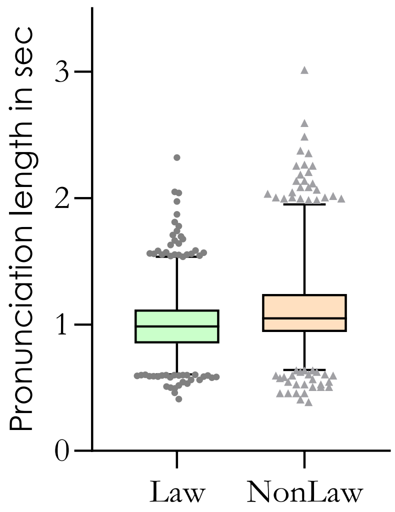


Figure 15. Differences between the two groups concerning the length of pronunciation

The considerable differences among the groups evidence the diversity in the pronunciation length of the participants. The average length of the pronunciation of the law group is 1.021317 seconds, while that of the non-law group is 1.228391 seconds during the first repetition. Both groups became faster during the second and the third repetition, but the difference between them remains statistically significant. Both groups contain outliers; however, there are more in the non-law groups. This indicates that some of the legal words required more effort from the participants than others. The following analysis was conducted to investigate whether it was familiarity with the item or separate factors that determined the length of the pronunciation.

In comparison with experiments that measure reaction times, there is no time limit for the length of pronunciation. A faster or slower pronunciation of the item might shed light on the underlying processes accompanying lexical retrieval and phonological encoding.

Although the outliers and influential data points were not removed from the dataset, the distribution of the length of pronunciation was inspected for each item and subject separately before fitting a model to the data. This was done with *quantile-quantile plots*, using the *qqmath()* function from the *lattice* package. Most of the subjects' length of pronunciation follows a normal distribution; only 2 out of the 32 participants have thick right tails. They will be inspected later to see whether they are influential data points for the model. However, only one of the words, *Strafvereitelung,* was an outlier.

In order to evaluate the effect of the subject's familiarity with the linguistic item and other factors on temporal reduction in the word naming task, a linear mixed-effects model was fitted to the data. All participants pronounced the words in the naming experiment three times. The first step in the analysis was to examine a control variable for possible longitudinal effects of familiarisation or fatigue. This was done using the *xylowess()* function from the *languageR* package. Unsurprisingly, the effect of the REPETITION was the same across the subjects: the participants' length of pronunciation decreased with the number of trials.

The linear mixed-effects model (Baayen et al. 2008) was fitted to the data using the function *lmer()* from the *lme4* package in the R software program (Bates et al. 2015). According to Baayen & Milin (2010), mixed-effects models obviate the necessity of prior averaging over participants and/or items and thereby offer the researcher the far more ambitious goal of modeling the individual response of a given participant to a given item. P-values for fixed effects were calculated by means of the R-package lmerTest (Kuznetsova et al. 2017).

Following the approach of McConnell & Blumenthal-Dramé (2019) baseline model was fitted to the pronunciation length as an initial step. This model contained WORDLENGTH as a fixed effect and a by-SUBJECT intercept as a random effect. The justification for fitting such a baseline model was to correct the response variable for effects that are known to modulate pronunciation length in the rapid naming experiments but are independent of the experimental manipulation (Linzen & Jaeger 2015). The per-word residual pronunciation lengths of the baseline model are the corrected pronunciation lengths that were used for further analysis. Two identical models were fitted to the data to examine the effect of the corrected pronunciation length on the automatic model selection. The first model used the raw pronunciation lengths as the dependent variable and the other, the corrected length. The two models converged in their results.

For model selection, the corrected length was used. In the initial phase, a model containing all the explanatory variables was fitted to the data, with ITEM as a random intercept. The model selection was performed in a backward manner; with each step, the explanatory variables with the highest p-value were removed, and the models were compared with the *anova()* function. From the social variables EXPOSURE ($\chi2$ = 1.1565, df = 1, p = 0.282) did not reach the significance level, AGE rendered a marginally significant effect (χ2 =3.2969, df = 1, p = 0.069) and OCCUPATION (χ2 =30.003, df = 1, p = 4.314e-08) rendered a significant effect on the rate of temporal reduction. Two of the frequency measures did not reach the significance level: FREQDWDS (χ2 =1.4084, df = 1, p = 0.2353), FREQBASE (χ2 =1.1286, df = 1, p = 0.2881). The rest of the linguistic variables have reached the significance level: FREQSTEMENDING (χ2 =0.205, df = 1, p = 0.6507), FAMILIARITY (χ2 =892.2, df = 1, p = 2.2e-16), REPETITION (χ2 =109.17, df = 1, p = 2.2e-16), FREQWEB (χ2 =12.377, df = 1, p = 0.0004). READING also turned out to be a significant predictor of temporal reduction (χ2 =153.74, df = 1, p = 2.2e-16).

After eliminating all the insignificant predictors, the interaction between the social and linguistic variables was explored. Interaction between OCCUPATION and GENDER turned out to be significant (χ2 =56.066, df = 1, p = 7.088e-14), just as the interaction between OCCUPATION and AGE (χ2 =7.5655, df = 1, p = 0.005). For this reason, a three-way interaction between OCCUPATION, AGE and GENDER (χ2 =316.76, df = 8, p = 2.2e-16) was added to the final model.

Besides the interaction between the social variables, the interaction between the linguistic variables was investigated. Firstly, a significant relationship was found between FREQWEB and z-scores of FAMILIARITY (χ2 =101.52, df = 4, p = 2.2e-16). Secondly, FAMILIARITY and TP_COMPOUND also reached significance. (χ2 =125.6, df = 4, p = 2.2e-16). The model with the three-way interaction between FREQWEB, FAMILIARITY, and TP_COMPOUND reached a higher significance than the model with the two-way interaction (χ2 =161.39, df = 8, p = 2.2e-16. The final model with the interactions is summarised in Table 9 below.

Table 9. Fixed effects of the linear mixed-effect model of best fit

	Estimate	**Std. error**	***t* Value**	***Pr* (> |*t*|)**
(Intercept)	9.169e-01	1.052e-01	8.717	2.82e-13
Gender:Male	-3.942e-02	1.311e-02	-3.007	0.002652
Occupation:Non	-2.058e-02	8.890e-03	-2.315	0.02066
Familiarity	-1.632e-02	1.534e-02	-3.860	0.000851
Repetition 2	-4.010e-02	5.131e-03	-7.815	6.49e-15
Repetition 3	-5.306e-02	5.131e-03	-10.337	< 2e-16
LogFreqWeb	-3.223e-02	1.009e-02	-3.195	0.002357
TP_Compound	-5.689e-02	5.181e-02	-3.098	0.00145
logReading	1.095e-01	9.291e-03	11.791	<2e-16
GenderMale:OccupationNon	3.989e-02	1.888e-02	2.113	0.034655
OccupationNon:AgeOld	1.331e-01	1.762e-02	7.553	4.92e-14
OccupationNon:AgeYoung	-8.947e-02	1.5703-02	-5.698	1.27e-08
GenderMale:AgeOld	1.267e-01	1.911e-02	6.631	3.64e-11
GenderMale:AgeYoung	-1.610e-01	1.928e-02	-8.348	<2e-16
Familiarity:TP_Compound	1.006e-01	5.278e-02	1.906	0.056

LogFreqWeb:TP_Compound	1.819e-03	5.565e-03	0.327	0.7438
Familiairity:LogFreqWeb	-8.077e-03	1.689e-03	-4.782	1.78e-06
GenderMale:OccupationNon:Age:Old	-9.812e-02	2.531e-02	-3.876	0.000107
GenderMale:OccupationNon:Age:Young	1.269e-01	2.782e-02	4.563	5.14e-06
Familiarity:LogFreqWeb:TP_Compound	-1.077e-02	5.768e-03	-0.431	0.011461

8.1.3 Experiment III

The dependent variables used for the statistical analysis stem from the sentence reading and recall task. The two parts of the experiment will be analysed separately. For each participant, the recordings of the sentence reading task were analysed to elicit information on temporal reduction. Different methods were implemented to automatise this process, including the WebMAUS Basic service of the Munich automated segmentation system hosted at the website of the Ludwig Maximilian University and Transcribear, a new online transcription and annotation tool supporting automated speech-to-text (developed by Yu-Hua Chen[86]). These software programmes varied in their accuracy in aligning the spoken text. Additionally, the sentences, bigrams, and nouns were manually corrected for length. For each participant, the length of the pronunciation of the sentence, the bigram and the selected nouns were measured. The length of the sentence was assumed to deliver information on the general speech rate of the participants; the length of the bigrams and nouns can reveal some information and processing and articulation speed that could be due to holistic storage. SPEECH RATE was coded as a continuous variable and was measured as the number of phonemes pronounced per second.

The independent variables used to predict the rate of temporal reduction are divided into three different groups, extralinguistic, linguistic and cognitive variables. The group of extralinguistic variables include AGE, GENDER, OCCUPATION, and hours spent watching CRIME SERIES in a usual week.

AGE of the participants ranged between 20 and 30 years, with a mean of 24. It is not expected to find any significant effects that AGE has on SPEECH RATE since all the participants are approximately the same age. The ageing of the voice and decline of memory capacities are observed in the older population. AGE was coded as a continuous variable.

[86] There is no publication available that could be cited. The software was advertised via the corpus list. It is available at https://transcribear.com/

GENDER has a controversial status; some researchers claim that there are differences between men and women in respect of rate of spectral and phonological reduction[87], holistic retrieval vs online assembling. Ullman and colleagues (2007) suggest that there are gender differences in a wide range of tasks involving language. Research has shown that episodic memory tasks and verbal fluency tasks revealed differences between the performance of male and female participants. The study of Ullman and colleagues (2007) has additionally demonstrated that women exhibit more potent effects of usage frequency and tend to memorize complex strings (e.g., *walked*), which men will instead compute compositionally (e.g., *walk* + *ed*). The study carried out by Blumenthal-Dramé (2012) produced similar results. Female participants had an advantage with respect to the efficiency of morphological decomposition. Blumenthal-Dramé (2012) explained this with the fact that females recruit additional brain areas handling the disentrenchment of complex strings. In experiment II, gender has also shown a significant impact on the rate of temporal reduction.

OCCUPATION is a binary variable representing the two groups of participants, law students and students outside the field. It is assumed that there is no difference between the performance of the two groups regarding everyday language stimuli. However, law students should have an advantage in processing and remembering the bigrams that are strongly associated with the field of law.

The last factor of this group is the time participants spend watching CRIME SERIES. This information was elicited from the background questionnaires. Participants were asked to indicate whether they watch crime series and how many hours they spend watching during an average week. Participants who watch crime series may be more likely to be exposed to law words. This prior exposure is assumed to be reflected both in their speech fluency and accuracy of recall.

The group of linguistic variables is the most numerous; it encompasses the following variables: ITEM TYPE, CONDITION, FAMILIARITY, FREQUENCY, TRANSITIONAL PROBABILITY (TP), MUTUAL INFORMATION, LOG-LIKELIHOOD, T-SCORE.

ITEM TYPE is a binary variable representing the two stimuli groups, legal and everyday bigrams. CONDITION is a threefold distinction among high, and low-transitional probability bigrams and frequent bigrams. High-transitional probability bigrams were expected to be pronounced with a faster speech rate and remembered more accurately than the other two groups.

[87] For discussion on the relationship between gender and speech rate see Yuan, Jiahong, et al. (2006), Byrd (1994), and on gender and spectral reduction Guy (1992), Raymond et al. (2006).

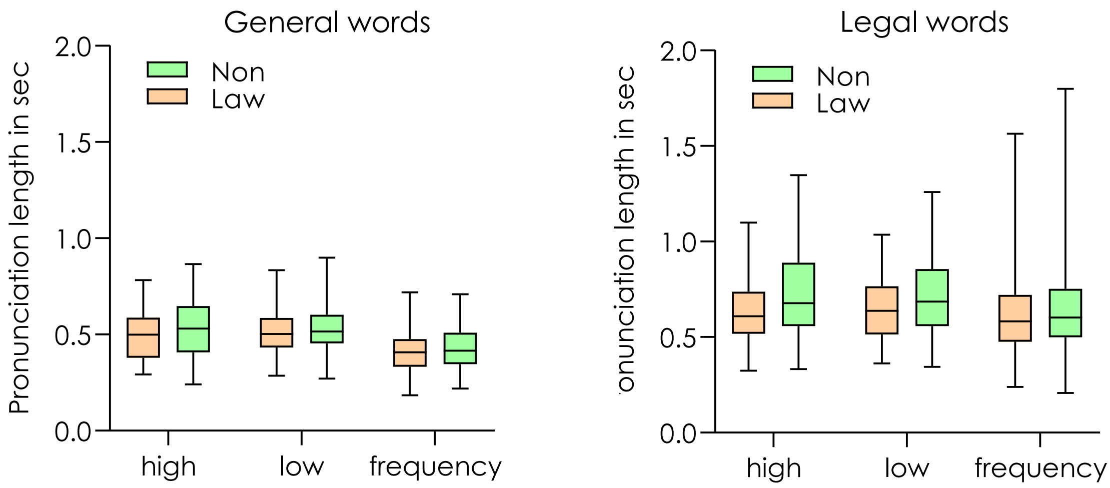


Figure 16. The effect of condition across groups and item types

Figure 16 reveals that high-transitional probability bigrams had no processing advantage compared to high-frequency bigrams.

As per the previous experiments, FAMILIARITY was obtained from the familiarity judgement task converted into Z-scores. High-frequency general words were judged as the most familiar items from the list, followed by general items belonging to the high-transitional probability group. A similar tendency is observed regarding the legal words; however, all items were rated as less familiar compared to the general words.

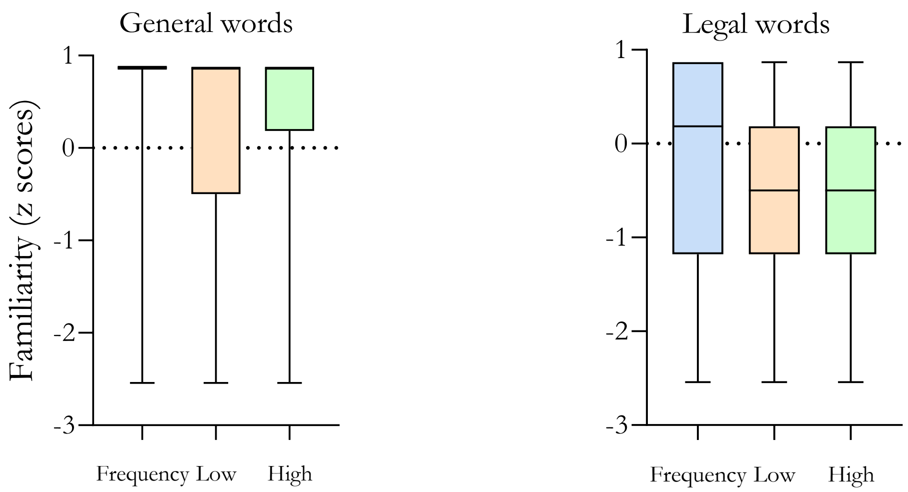


Figure 17. Familiarity scores in relation to item types and conditions

FREQUENCY is the most repeatedly cited factor in connection with entrenchment, particularly with reduction and chunking. Previous experiments in this study showed that FREQUENCY is a reliable, significant predictor. For the present experiment, frequency information was obtained for the nouns (FREQNOUN), modifiers (FREQMODIFIER) and bigrams (FREQCOLLOC) from five different corpora, namely GermanWeb, Kernkorpus 1900-1999, Kernkorpus 2000-2010, Die ZEIT, Filmuntertitel. These corpora include a wide range of genres and text types. The selected nouns are of different frequencies, representing low, medium, and high-frequency words, following the threshold levels established by Francis and Kucera[88]. The frequencies from the different corpora strongly correlate. This implies that if frequency measures show a significant effect on speech fluency or accuracy, then this is not due to one particular corpus but because the selected bigrams are entrenched in the mental lexicon of the participants. Prior exposure to these measures determines the speed of access and the strength of memory representation. The correlation between the frequency of the bigrams in the different corpora is illustrated in Figure 18. FREQUENCY was log-transformed before the statistical analysis and coded as a continuous variable. The frequencies obtained from the GermanWeb were used for the statistical modelling because this is the largest among the used corpora.

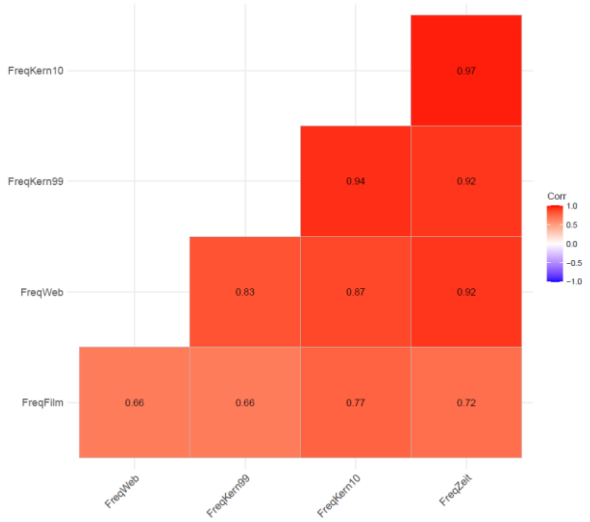


Figure 18. Correlation between the frequency measures obtained from different corpora

The other explanatory variables within the group of linguistic factors focus on the association strength between the words. These include TRANSITIONAL PROBABILITY (TP), MUTUAL INFORMATION, LOG-LIKELIHOOD, T-SCORE.

[88] For more details, see Burgess & Livesay (1998)

TP, as was mentioned in 3.2.2, was calculated using the method developed by Jurafsky et al. (2001a). It was measured for all the bigrams in all five corpora. The different corpora produced different rankings, which suggests that it is possible that transitional probability measured in corpus A might be a better predictor of the participants' performance than corpus B or C because it reflects the mental representation of the bigrams more accurately. This difference, if it exists, can be assumed to be connected to text types and genres that are the backbone of the corpus, which indirectly can tell us about the prior exposure to selected bigrams.

The remaining association measures are mathematical formulae that interpret co-occurrence frequency data. For each bigram extracted from a corpus, an association score was computed that indicated the amount of (statistical) association between the two words. The co-occurrence data for the bigrams were organized in contingency tables following the method of Evert & Krenn (2001). The contingency tables[89] included the observed and expected frequencies of the bigrams in the selected corpora.

MUTUAL INFORMATION (MI) is a measure used to determine the co-occurrence strength between two words. A high MI score indicates a frequently co-occurring word pair. It is calculated using the information from the contingency table. The formula is the following: $MI = log\frac{O_{11}}{E_{11}}$, where O stands for observed, and E for expected frequency of the bigram in the respected corpus.

LOG-LIKELIHOOD (LL) is a well-accepted measure of statistical significance; the higher the score, the more likely it is that the association is not due to chance. However, more frequent words tend to have higher log-likelihood scores because there is more evidence for such words. The formula that was used to calculate log-likelihood is the following: $log - likelihood = 2\sum_{ij} O_{ij}\, log\frac{O_{ij}}{E_{ij}}$. The size of the corpus highly influences the log-likelihood score; thus, the direct comparison of the log-likelihood scores among the different corpora is not possible. T-SCORE is a measure not of the strength of association but the confidence with which one can assert that there is an association between the two words. The formula to calculate it is the following: $t - score = \frac{O_{11}-E_{11}}{\sqrt{O_{11}}}$. Similarly, T-SCORE is also dependent upon the corpus size. In addition to the above-described association measures, delta P and Pearson's chi-squared were calculated for the bigrams.

[89] For an excellent description of contingency tables, see Evert (2009) section 5.1

As figure 19 reveals, a strong positive Spearman correlation holds between almost all association measures, with the strongest correlation between delta P and TRANSITIONAL PROBABILITY, T-SCORE and LOG-LIKELIHOOD. This strong correlation does not allow a direct comparison between the different association measures, as it would result in multi-collinearity in a statistical model. For this reason, DELTA P and T-SCORE were excluded from the analysis.

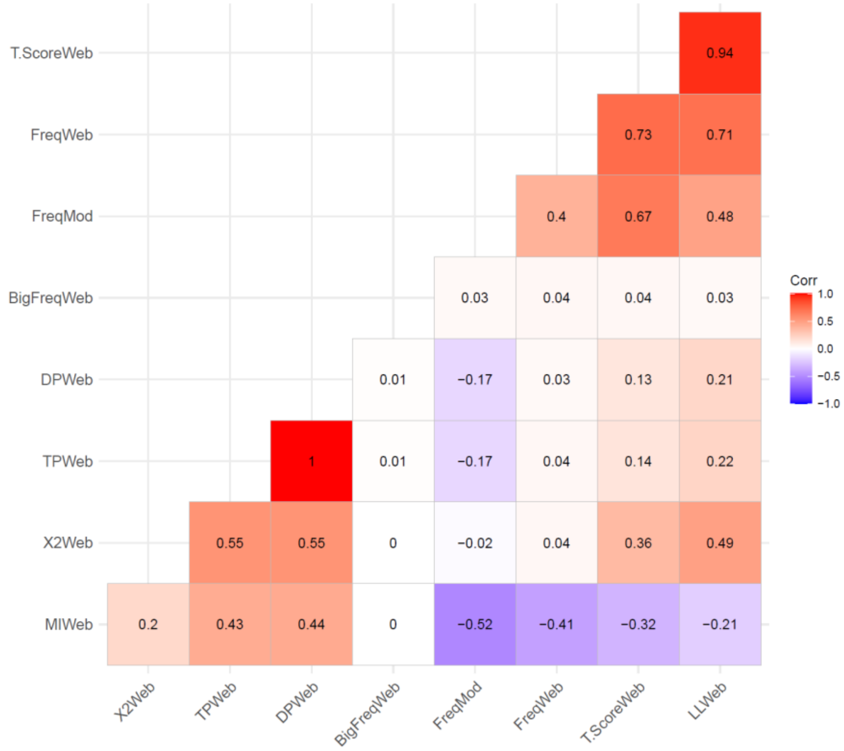


Figure 19. Correlation between different association measures

The last group of explanatory variables is aimed at measuring the individual difference among the participants and includes the results of the different memory tests and the reading habits of the subjects.

READING is collected during the self-paced reading task. It includes the reading times of the law and general texts. As the time participants spent answering the questions was inflated due to potential distractions, only the actual reading time (in seconds) was included. It is coded as a continuous variable.

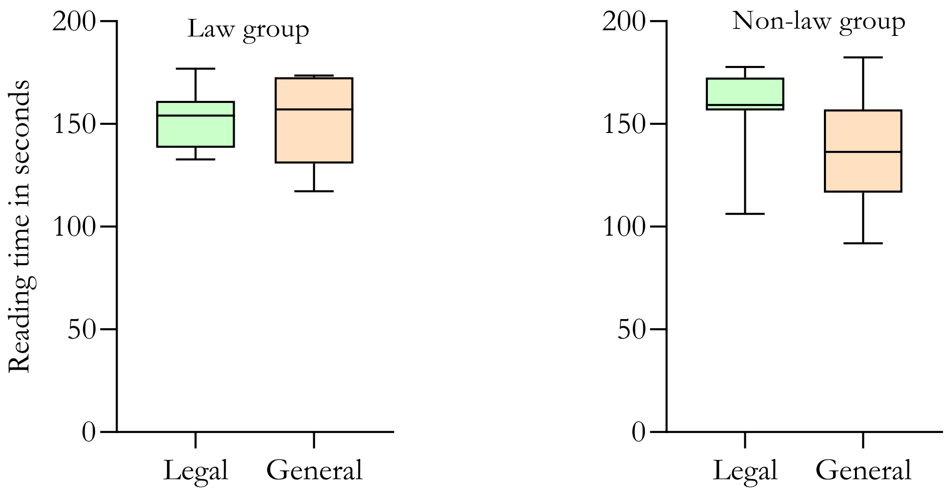


Figure 20. Reading rates of the law and non-law groups in experiment III

Figure 20 reveals that the tendency observed in experiment II also continues in experiment III. Participants that belong to the law group processed the legal texts faster than the general texts. An interesting observation is that members of the non-law group process the general text faster than law group members. It is very unlikely that the difference is due to age or educational differences between the participants, as both these factors were controlled for.

READING ACCURACY is part of the reading comprehension task, and it measures the number of correct answers participants have provided to the comprehension questions. The different text types are calculated separately. Each text type included 10 comprehension questions.

READING SPAN is an explanatory variable that measures the memory span of the participants during the reading span task. The author[90] of the reading span task provides an open-source R-script[91] to analyse the results. The script offers different score types. For the statistical analysis, I opted for the prop. score, which is calculated using the proportion of correct answers. The maximum achievable score was 20, and the participants' performance varied from 3-17, with a mean of 9,64 points. Moreover, there was no correlation detected between the participants' age and their performance of the reading span task.

[90] For the full description of the task see Stone & Towse (2015)

[91] The script is available at http://www.cognitivetools.uk/cognition/tasks/Verbal-WM/readingSpan/

WORD SPAN describes the percentage of correctly recalled words. Unlike the established way of calculating the score, where the last item size recalled correctly is taken as a measure of word span, the overall percentage of correctly recalled items was used for the analysis. The reason for deviating from the traditional way of calculating the word span was that almost every participant could recall seven words. In order to measure the differences between the subjects, each item size was first analysed separately, and then the results were averaged. Conwey et al. (2005) also point out a problem associated with the standard score-calculating methodology: the difficulty of a span item may vary on many dimensions, thus threatening span reliability across different tasks (or different versions of the same task). The dimension includes the length of the word, the semantic similarity of the stimuli within one item size, or the display duration of the word. They argue that information on all other trials is discarded by simply estimating the item size at which a subject falls below a given threshold (and then ending the task). The threshold usually varies between 2 and 6, which significantly limits the sensitivity of the measure. Thus they claim that absolute span scores are inappropriate for individual-differences research (2005: 774). The participant's performance varied from 71% to 97%, with an average of 83% correctly recalled words. Unlike in the case of READING SPAN task, the participants' age influenced their scores in the WORD SPAN task. There is a moderate positive but significant correlation between the two factors ($R=0.38$, $p<0.001$).

Mental processing speed measures how quickly (MENTPROCRT), and accurately (MENTPROCACC) subjects process the given information and make a certain decision based on it. The measure includes two separate values: speed and accuracy. Due to the difficulty of finding an established and accepted score-calculating methodology, speed and accuracy are two explanatory variables. Speed is measured in the form of reaction times, which were log-transformed for the statistical analysis. Accuracy is the percentage of correctly performed tasks. Contrary to the word span task, there is a weak negative but significant correlation between the participants' age and their MENTPROCRT ($R=0.22$, $p<0.001$).

8.1.3.1 Modelling temporal reduction

The above-described tasks aimed to compare the predictive power of FAMILIARITY, FREQUENCY and four association measures. Moreover, they aimed to explore the effect of CONDITION (high-, low-transitional probability and high-frequency bigrams) and reading habits on the rate of temporal reduction.

In order to assess how SPEECH RATE is conditioned, a mixed-effects model was fitted for each, with ITEM used as random slopes (Baayen et al. 2008). Due to a large number of explanatory variables and a rather small data set, the modelling was done first separately for the linguistic, cognitive, and social factors. The models' dependent variable was the raw pronunciation length of the nouns.

The discussion will focus initially on linguistic factors. The correlation matrix between the various linguistic factors showed a high correlation among transitional probability, mutual information, chi-squared, log-likelihood and t-score. This high correlation made it impossible to include all the linguistic variables into one global model. In the initial step, the variables were chosen that did not show correlation; included FREQNOUN, FREQMODIFIER, FREQCOLLOC, TP, ITEM TYPE, CONDITION, and FAMILIARITY. Four of these linguistic predictors reached the significance level (FREQNOUN ($\chi2$ = 22.35, df = 1, p = 2.272e-06), FREQMODIFIER ($\chi2$ = 6.0076, df = 1, p = 0.014), ITEM TYPE ($\chi2$ = 4.5566, df = 1, p = 0.0347), FAMILIARITY ($\chi2$ = 8.5869, df = 1, p = 0.0033)). Using the backward selection method, the predictors with the highest p-value were removed one by one. After eliminating predictors, the models were compared using the *anova()* function. As the next step, possible interactions between the linguistic predictors were investigated. The only interaction that reached a marginal significance (p-value = 0.07082) was between TRANSITIONAL PROBABILITY and ITEM TYPE.

Because of the high correlation between the corpus-based association measures, separate models were fitted using these explanatory variables, and then the resulting models were compared. Replacing FREQMODIFIER with LOGLIKELIHOOD resulted in the same AIC value and explanatory power. All the other association measures failed to reach the significance threshold.

The same procedure was applied to the cognitive and social factors. From the social factors, two variables reached significance, CRIME SERIES ($\chi2$ = 46.033, df = 1, p = 1.163e-11), OCCUPATION ($\chi2$ = 44.609, df = 1, p = 2.406e-11). Out of the 5 cognitive variables, 3 rendered a significant effect on the rate of temporal reduction: MENTPROCRT ($\chi2$ = 16.569, df = 1, p = 4.69e-05) MENTPROCACC ($\chi2$ = 8.9339, df = 1, p = 0.0027), and WORDSPAN ($\chi2$ = 19.172, df = 1, p = 1.194e-05). After selecting all the significant predictors from the three groups, they were merged into one global modal. All the previously selected predictors remained significant in the merged global model. This final model is summarised in Table 10.

Table 10. Summary of the final model on temporal reduction in Experiment III

	Estimate	**Std. error**	***t* Value**	***Pr* (> |*t*|)**
(Intercept)	1.236e+00	1.332e-01	9.274	6.76e-13
FreqNoun	-4.554e-02	9.225e-03	-4.937	6.80e-06
FreqModifier	-8.091e-03	3.459e-03	-2.339	0.01963
Familiarity	-1.358e-02	4.431e-03	-3.066	0.00222
ItemType:Legal	8.942e-02	3.421e-02	2.614	0.01121
TP	-1.595e-04	2.841e-04	-0.561	0.57462
CrimeSeries	-1.468e-02	2.152e-03	-6.821	1.48e-11
Occupation:Non	5.828e-02	8.673e-03	6.719	2.92e-11
MentalProcessing(RT)	-1.427e-02	3.509e-03	-4.067	5.10e-05
MentalProcessing(Acc)	-1.010e-02	3.386e-03	-2.984	0.00291
WordSpan	-2.187e-02	4.995e-03	-4.377	1.31e-05
ItemType:Legal*TP	-1.966e-03	1.105e-03	-1.779	0.07558

To resolve the issue regarding different measurements of speech rate (raw pronunciation length, phon/sec, syll/sec), the final model's dependent variable was first replaced by residualised pronunciation length, then by the phon/sec measure, and finally by syll/sec speech rate. The residualised pronunciation length was extracted from a baseline model. This model contained only WordLength as fixed effects, and a by-Subject and by -Item intercepts as random effects. The different approaches to the same matter produced mixed results. Unexpectedly the models did not converge. The comparison of the different measures is provided in Table 11 below.

Table 11. Comparison of the different speech rate measures

	Residualised PronLength	**Syll/sec**	**Phon/sec**
(Intercept)			***
FreqNoun			**
FreqModifier	*		**
Familiarity			
ItemType:Legal			
TP			
CrimeSeries		***	***
Group:Non		***	***
MentalProcessing(RT)		***	***
MentalProcessing(Acc)		**	**
WordSpan		***	***
ItemType:Legal*TP			
Familiarity*GroupNon	**		

Table 11 reveals that the phon/sec measure is the closest to the pronunciation length measure. Residualising the pronunciation length removes the significance of most of the explanatory variables. Interestingly, only the cognitive and social predictors are significant in the model that uses the syll/sec measure. Due to these differences between the models, WORDLENGTH was added to the model. This allows for controlling its effect on temporal reduction statistically. Adding this predictor to the final model increased the model fit ($\chi2$ = 44.458, df = 1, p = 2.599e-11). Moreover, WORDLENGTH alone explains 14% of the variance (R^2 = 42.3 for the model without WORDLENGTH, and R^2 = 56.4 with WORDLENGTH). The comparison of the different speech rate measures showed that the methods of calculating the speech rate influenced the outcome of the selection of the mixed-effects regression model. These mixed results could be due to a large number of explanatory variables and the small dataset.

Furthermore, it was analysed whether there is a correlation between the average SENTRATEPHON[92] and the pronunciation rate of the selected noun SPEECH RATE. As phon/sec measure of speech rate turned out to be the most reliable, this measurement was used to calculate the speech rate for each sentence and participant.

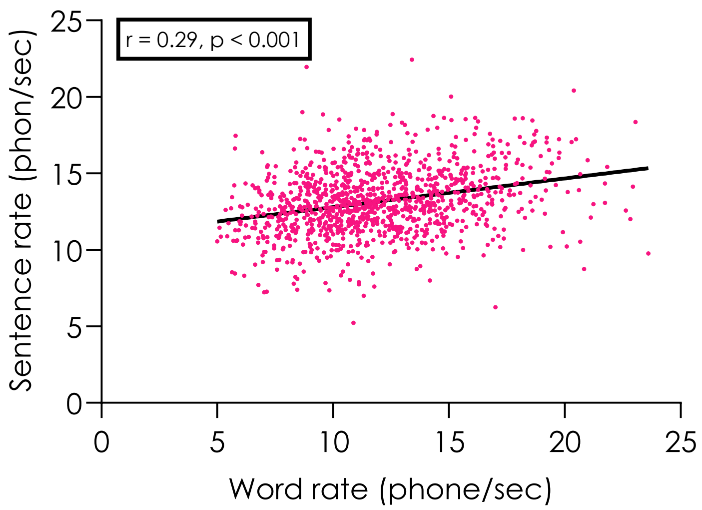


Table 12. Correlation between noun pronunciation length and sentence speech rate

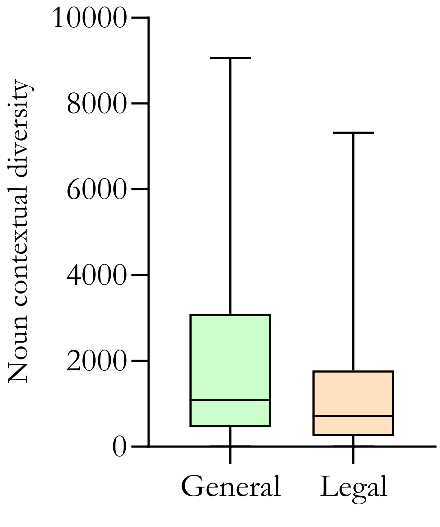
Due to the significant correlation between the two variables, SENTRATEPHON was added to the final model. The model with SENTRATEPHON is more significant than the previous model without it ($\chi2$ = 6605.7, df = 1, p = 2.2e-16), and this variable helped explain an additional 4% of the variance (R2 = 60). However, adding this variable to the model turned WORDSPAN from a significant to an insignificant predictor of pronunciation length

[92] Refers to the average speech rate of the participants during the sentence reading. It is measured in phon/sec.

(p=0.49). Once again, it has to be emphasised that the instability of the explanatory variables could be due to the small number of observations.

8.1.4 Experiment IV

Experiment III is a pilot study for experiment IV. For this reason, only the explanatory variables that differ between the two experiments will be highlighted here. The explanatory variables fall into social, linguistic, and cognitive categories, just as in the previous experiment. The social variables are the same except for the CRIME SERIES, which was not taken into account in this follow-up study. The linguistic factors were also slightly modified, the ASSOCIATION MEASURES that were derived from the corpora were removed, and TRANSITIONAL PROBABILITY was split into TP_FORWARD and TP_BACKWARD. TP_BACKWARD differs from TP_FORWARD in the fact that it measures how likely a word is preceded by another word. Moreover, CONTEXTUAL DIVERSITY was added to the list of linguistic predictors. It was measured as the number of films in which the word occurs in the subtitle corpus (Baayen 2010). It was obtained using the Filmutertitel corpus hosted on the website of DWDS[93]. The query used to extract CONTEXTUAL DIVERSITY was *count("Lemma" #within file)*. The CONTEXTUAL DIVERSITY for the nouns and the adjective-noun modifiers were calculated separately. The CONTEXTUAL DIVERSITY of the collocation was relatively low, varying between 1 and 322, with a mean of 13. The nouns provided more useful data than the collocations. The summary of the nouns CONTEXTUAL DIVERSITY (NCD) is illustrated in Figure 21.



[93] DWDS stands for the Digitales Wörterbuch des Deutsche Sprache. The Filmutertitel corpus includes 12008 movies that add up to 75605122 tokens.

Figure 21. Contextual diversity of the target nouns

Cognitive factors were altered for this group. The only measure that is common between the previous and this experiment is READING. In this experiment, it was reclassified as VERBALPROCESSING speed, but it was still measured as the amount of time spent on reading the text in milliseconds. The reason for reclassification was that the speed with which participants read is more likely to reflect their verbal processing speed and capacity than their reading habits. Participants were asked to read out sentences during the sentence reading and recall experiment. It is hypothesised that the speed of decoding during reading aloud and in silence is the same; therefore, this variable aims to explore the speed with which participants can decode information.
Moreover, the NONVERBALPROCESSING speed of the participant was added as a possible explanatory variable of temporal reduction. This measure was extracted from the task-switching experiment. It is presumed that some participants might be better at verbal processing while others at non-verbal processing, and these skills might compensate for each other.

8.1.4.1 Modelling temporal reduction

The modelling of temporal reduction followed the same steps as in all the previous experiments. The initial model included all the explanatory values, and in a backwards selection manner, all the insignificant variables were removed, and the models were compared using the function *anova()*.
The dependent variable was the speed of pronunciation of the target nouns in phon/sec. This measure was used as it considers the length of the word to a certain degree and is precise enough to detect individual differences between participants and between items. Item was treated as a random effect. Frequency measures were log-transformed and scaled; all the other reaction time data were also scaled. A correlation test was performed to ensure that model results are not biased due to the correlation between the linguistic predictors. This correlation is illustrated in Figure 22 below.

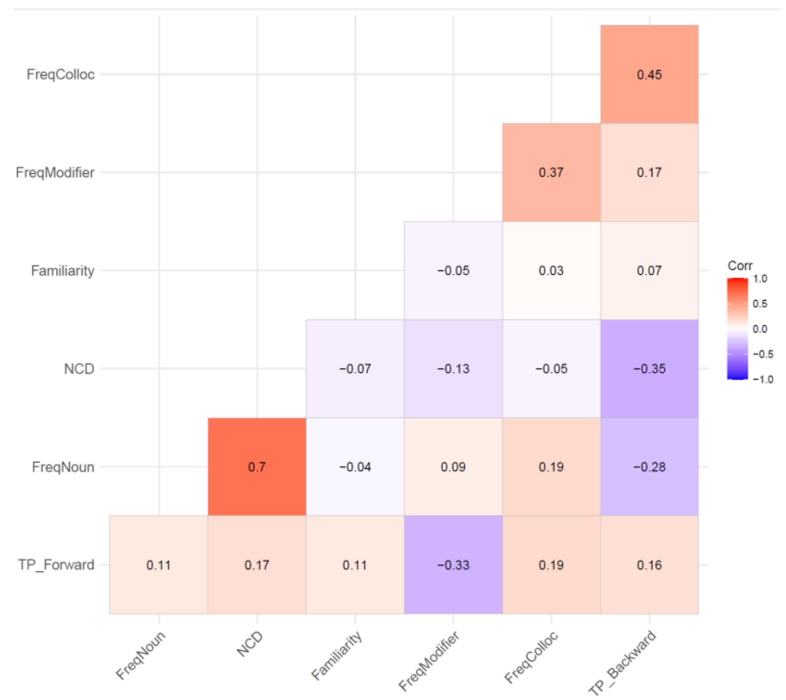


Figure 22. Correlation between the linguistic variables

As Figure 22 illustrates, only the correlation between FREQNOUN and the NCD (noun contextual diversity) is too high to include them in the same model. The other two moderate correlations that appear high in the diagram were included in the initial global model because their VIF value was below the 5[94] (LOGFREQMOD=4.44 and LOGFREQCOLL=3.86).

None of the social factors rendered a significant effect on the rate of temporal reduction: GENDER ($\chi2$ = 0.2088, *df* = 1, *p* = 0.6477), OCCUPATION ($\chi2$ = 0.5423, *df* = 1, *p* = 0.4615), AGE ($\chi2$ = 2.5716, *df* = 1, *p* = 0.1088). Out of the linguistic variables three reached significance: FAMILIARITY ($\chi2$ = 242.47, *df* = 1, *p* = 2.2e-16), TP_FORWARD ($\chi2$ = 21.358, *df* = 1, *p* = 3.811e-06), and FREQCOLLOC ($\chi2$ = 3.5451, *df* = 1, *p* = 0.05). The remaining linguistic variables did not render a significant effect: FREQMODIFIER ($\chi2$ = 3.345, *df* = 1, *p* = 0.187), FREQNOUN ($\chi2$ = 3.2957, *df* = 1, *p* = 0.129), TP_BACKWARD ($\chi2$ = 1.471, *df* = 1, *p* = 0.225). Both variables from the group of cognitive variables have reached the significance level: VERBALPROCESSING ($\chi2$ = 4.4354, *df* = 1, *p* = 0.0035), NONVERBALPROCESSING ($\chi2$ = 22.944, *df* = 1, *p* = 1.668e-06).

[94] VIF (Variable Inflation Factors) determines the strength of the correlation between the independent variables. It is predicted by taking a variable and regressing it against every other variable. VIF score of an independent variable represents how well the variable is explained by other independent variables. Levshina (2015) suggest 5 as the threshold level below which correlation between the explanatory variables does not affect the model.

After removing all the insignificant explanatory variables, the interaction between the linguistic predictors was investigated. One significant and two marginally significant interactions were detected among the linguistic variables. A significant interaction is observed between FAMILIARITY*FREQCOLLOC.

The marginally significant interactions are between TP-FORWARD*ITEM TYPE and among TP_FORWARD * FAMILIARITY *FREQCOLLOC.

Table 13 provides a summary of the final model and all interactions. Despite the marginal significance of the interaction, the model with the interaction is more significant than the one without it ($X2$ =16.442, df = 6, p =0.01157).

Table 13. Summary of the final model of temporal reduction experiment IV

	Estimate	Std. error	*t* Value	*Pr* (> |*t*|)
(Intercept)	7.466e+00	3.781e-01	19.748	< 2e-16
WordLength	3.958e-01	4.429e-02	8.937	2.54e-15
TP_Forward	4.244e-01	1.747e-01	2.430	0.01626
Familiarity	4.052e-01	2.561e-02	15.821	< 2e-16
ItemType:Terminology	-1.379e-03	3.215e-01	-0.004	0.99659
NonVerbalProcessing	-1.500e-01	3.142e-02	-4.775	1.84e-06
VerbalProcessing	-6.830e-02	3.146e-02	-2.171	0.02996
FreqColl	2.153e-01	1.440e-01	1.496	0.13780
TP_Forward*ItemType:Term	6.345e-01	3.622e-01	1.752	0.08278
TP_Forward*Familiarity	2.317e-02	2.773e-02	0.835	0.40356
Familiarity*FreqColloc	-7.782e-02	2.520e-02	-3.088	0.00202
TP_Forward*FreqColloc	5.170e-02	1.405e-01	0.368	0.71351
TP_Forward*Familiarity*FreqColl	-4.807e-02	2.586e-02	-1.859	0.06312

8.2 Phonetic/segmental reduction

Phonetic reduction has been analysed in the two main studies, experiments II and IV. The reduction analysis was restricted to the prenasal schwa elision in experiments II and IV and intervocalic /t/ elision in experiment II. Most of the explanatory variables were taken from the modelling of temporal reduction. Additionally, information about the phonetic environment was added to the modelling of phonetic reduction.

8.2.1 Experiment II

8.2.1.1 Prenasal schwa elision[95] in the selected words

The following experiment investigates the pre-nasal schwa deletion. Vowel reduction can be caused by several different factors, including an unstressed position at lower

[95] Elision and deletion are used as synonyms.

prosodic boundaries and increased speech tempo, general durational variation in connected speech, deaccentuation in function words as compared to lexical words, and in citation forms as compared to clear speech (Mooshammer & Geng 2008)Kohler (1977) explains schwa elision as the result of the time-stressed nature of German. The elision of schwa occurs most frequently in a pre-nasal position. In a later article (Kohler 1991), he specifies that shwa elision in word-final position is a regular phenomenon. According to Grantham O'Brian & Fagan (2016), there are cases when the schwa is not deleted before an /n/. The sound that precedes the schwa must be an obstruent (a plosive or a fricative). Additionally, the preceding sound cannot be a nasal or an approximant. Finally, if the schwa is in the diminutive suffix, then it cannot be deleted either.

The sound following the schwa should be in the coda of the syllable (Grantham O'Brian & Fagan 2016). All the selected words contain a schwa before an /n/. However, the word *Zwangsmaßnahmen* does not meet the requirements mentioned above for the schwa deletion, and therefore it will be omitted from the analysis.

Before determining the presence or absence of the shwa, one has to understand the phonetics of the schwa vowel. According to the International Phonetic Alphabet, schwa is characterised as a mid-central vowel. Following both English and German phonology rules, the shwa usually occurs in an unstressed syllable. This fact led to the assumption that vowel reduction involves approximation to the centre of the vowel space (Flemming 2009). The fact that schwa occurs in an unstressed position creates a favourable environment for elision. It is known that there is a strong correlation between word stress and vowel reduction. Unstressed syllables are less prominent and undergo more frequently phonological reduction processes. Another feature of schwa is that it tends to assimilate strongly to the surrounding sounds, which results in a substantial variation in the vowel quality of the sound. This is referred to as the coarticulation effect (Mooshammer & Geng 2008). The presence or the absence of the schwa was determined with the help of the spectrogram of the recordings and the information contained in the TextGrid file concerning the length of the schwa. The TextGrid file was obtained from the Munich Automatic Segmentation platform, using the WebMOUSE basic application. The presence of the schwa on the spectrogram was determined using the formants. If all three formants were visible on the spectrogram and the WebMOUSE assigned the schwa temporal characteristics, it was considered present in the word and the other way around, respectively.

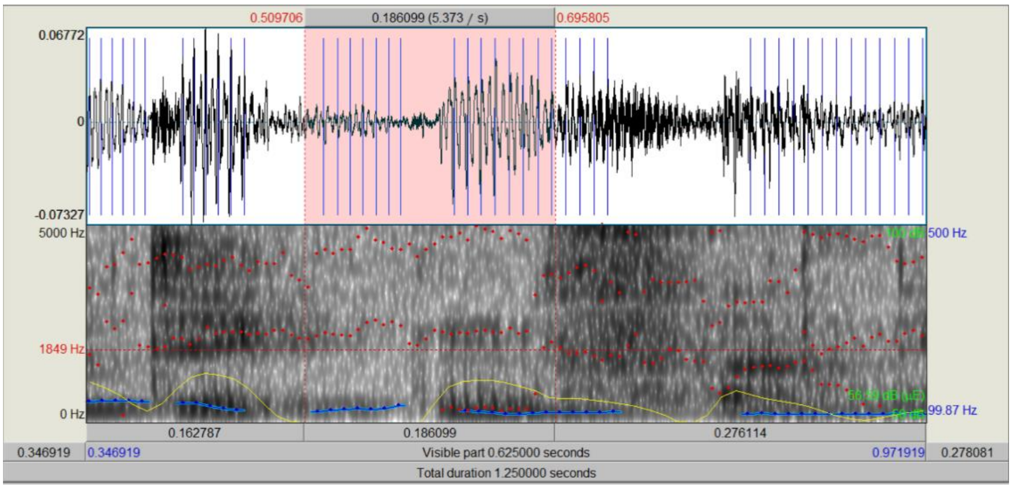


Figure 23. The word *Beschuldigtenvernehmung* without a pronounced schwa. The syllable -ten is marked with light red

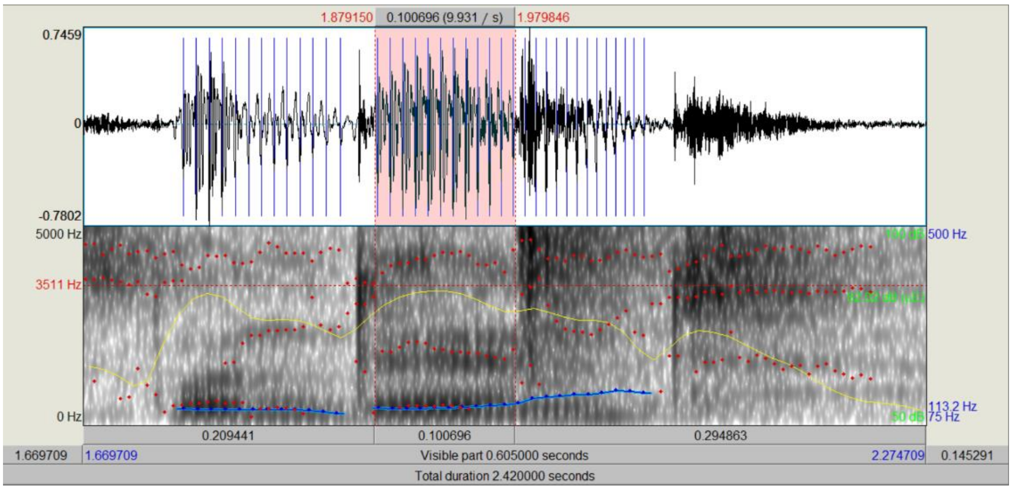


Figure 24. *Beschuldigtenvernehmung* with a pronounced schwa. The syllable -ten is marked with light red

Figures 23 and 24 show two cases, one containing the schwa and the other one demonstrating the schwa elision. The relevant sections are marked with light red (the syllable -*ten*). The formants of the German schwa are equidistantly distributed; the distance between them is approximately 1000 Hz (Machelett 1996). In Figure 17, the F1 and F2 are clearly visible, and the distance between them is approximately 1000 Hz. However, according to the theory, F3 is not as prominent as it should be. In Figure 16, on the hand, there is only a faint F1. The absence of the formants is the indicator of the deleted shwa.

The factor that made the analysis more complicated is that the schwa in all words is followed by a nasal, namely by /n/. The representation of an /n/ and the other nasals look like a faint vowel without much amplitude in the higher frequencies. Consequently, the sounds might look similar in the spectrogram.

The analysis of the schwa elision can be viewed from two different angles: a) by analysing the extralinguistic factors, i.e., why only certain participants have shown the elision of the schwa, and b) by analysing the phonetic features and lexical variables to establish why this particular word was phonetically reduced. The dependent variable for the analysis was the presence or absence of the schwa sound, and the explanatory variables were taken from the previous analysis.

The schwa elision is extensively studied in English[96], but less research is done in German. Two main corpora studies have investigated schwa deletion in English using corpus data. These were carried out by Dalby (1986) and Patterson et al. (2003). Dalby's (1986) research has shown that schwa elision occurs more frequently in fast-read speech (44% of the shwas were deleted) than in slow-read speech (2%). The study by Patterson et al. (2003) is based on the Switchboard corpus. The study concluded that the frequency of the word was directly correlated with the rate of shwa deletion. 15.4% of the high-frequency words demonstrated the shwa deletion, while only 6.2% of the low-frequency words.

The extra-linguistic variables in the present study include the following factors: AGE, GENDER, OCCUPATION and SPEECH RATE (in this case, the speed of pronunciation of the words under investigation measured in phon/sec). Given the mixed results from the previous, the current project will include both AGE and GENDER as explanatory variables. AGE was coded as a continuous variable. GENDER was coded as a binary variable (male, female).

To the best of my knowledge, OCCUPATION has not been explored as an explanatory factor regarding the rate of phonetic reduction. Gardner et al. (1987) analysed word frequency and lexical decision by examining lawyers, nurses and engineers. Their results indicated that error rates were higher when the participants judged unknown or unfamiliar words. This study will regard OCCUPATION as a possible explanatory variable that will forecast schwa elision.

SPEECH RATE is known to affect the reduction processes, including the word-final and internal /t,d/ deletion. Byrd & Tan (1996) concluded that a faster speech rate is associated with higher deletion rates, segment shortening and gestural overlap. These results were confirmed in a study by Raymond et al. (2006), where it was found that a

[96] For a detailed description of schwa elision in fast speech see Davidson (2006).

strong predictor of word-internal /t,d/ deletion. The present study considers the length of each word's pronunciation. It is coded as a continuous variable.

The lexical factors that might affect the spectral reduction phenomenon include word FREQUENCY, FREQSCHWAWORD, FREQSECMORPH, RELFREQBASE, TP_COMPOUND, FAMILIARITY, ITEM TYPE, STRESS PATTERN, REPETITION, WORD LENGTH. Raymond et al. (2006) have taken into account word form as a predictor of /t/ elision. The present study includes almost exclusively compounds.

FREQUENCY was coded as continuous variables for each token in the dataset from GermanWeb corpus. Besides extracting the frequency of the compound, the frequency of the word that contains the schwa (FREQSCHWAWORD), the frequency of the second morpheme (FREQSECMORPH)[97]. Moreover, as in the modelling of temporal reduction, the RELFREQBASE and TP_COMPOUND were added as possible explanatory variables. This corpus was chosen because it is a general corpus, and half of the participants were not related to the field of law, although the law corpus could be a better indicator of the performance of the lawyers. The raw frequency of the words was used, which was later log-transformed. Frequency and its effect on phonological reduction is a well-researched topic. Among the researchers who have established the role of frequency in language perception and production in general, in spectral phonological reduction, in particular, are Bybee (2002b, 2007b, 2010), Dell (1990), Ellis (2002, 2016), Schmid (2017b; 2010), Behrens & Pfänder (2016a).

EACH PARTICIPANT'S FAMILIARITY WITH THE WORDS WAS ELICITED and transformed into Z scores. ITEM TYPE as an explanatory variable is the simple distinction between words belonging to the legal sphere and general words. It was coded as a binary categorical variable (terminology, general). 13 of the selected words belong to the group of legal words and 7 to the general word group. The unequal division is the result of the restricted words available for the analysis.

The POSITION of the schwa sound within the word is regarded as a possible factor that determines the shwa elision. None of the selected words contains a pre-nasal schwa in the initial position. This factor was coded as a binary categorical variable indicating schwas in word middle and final positions.

The last factor, which is in the group of lexical variables, is REPETITION. The variable was coded as categorical (first, second, third). Each word was pronounced three times.

[97] The frequency of the compound Ermittlungsverfahren is 49840 (FREQUENCY), of the schwa word -verfahren is 1400170 (FREQSCHWAWORD), and of the second morpheme *Ermittlung* is 427266 (FREQSECMORPH).

Additionally, WORD LENGTH is included in the analysis. There is a direct correlation between the word's length and the pronunciation's duration. Dalby (1986) and Patterson et al. (2003) have observed that word length also affects schwa deletion. They found that the overall deletion rate was higher for three-syllable words than for two-syllable words.

The last aspect that is examined in respect of phonetic reduction is the phonological features of the words. This aspect includes the following variables: the PRECEDING AND FOLLOWING SEGMENTAL CONTEXT AND PHONEME BIGRAMS.

The PREVIOUS and FOLLOWING SEGMENTAL CONTEXTS describe the place of articulation of the sounds surrounding the target sound. Because this study investigates the schwa deletion in a pre-nasal position, only the previous segmental environment is regarded as a possible predictor of the rate of elision. The information concerning the previous sound is coded as a categorical variable (plosive, fricative, nasal, taps). The preceding sound's role in the schwa reduction is visualised below in Figure 25.

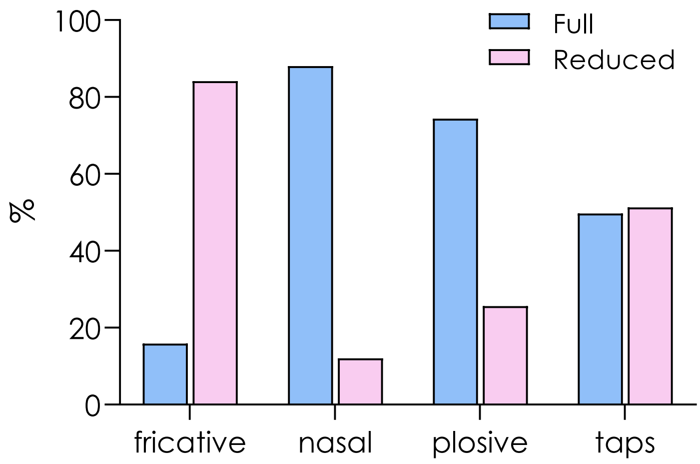


Figure 25. Prenasal schwa reduction and preceding sound

8.2.1.1.1 Modelling phonetic reduction

Before fitting the data to a generalised mixed-effects model using the *lme4* package, the raw frequencies from the corpus were transformed into log values, and additionally, all the continuous variables were scaled with the help of the *scale* function. The correlation between the frequency measures was determined to ensure that they would not lead to multicollinearity in the regression model. There was no significant correlation between the different frequency measures. The results of the correlation test are represented in Figure 26.

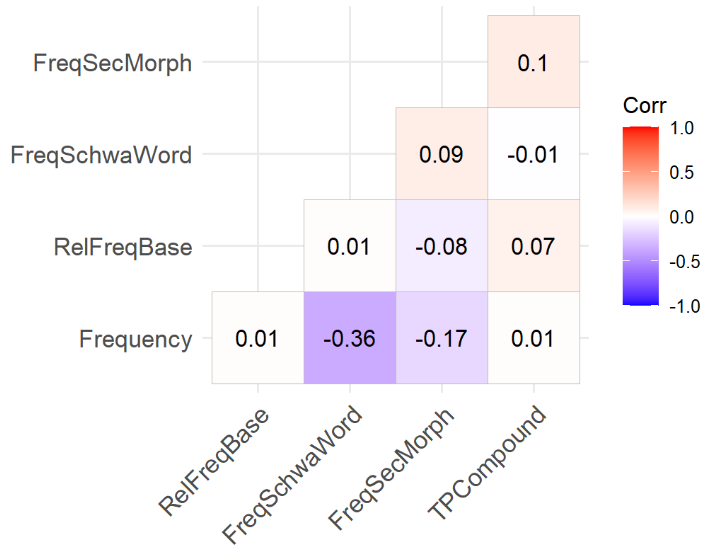


Figure 26. Correlation between the different frequency measures

The backward stepwise selection method was used to determine which dependent variables were significant for the schwa deletion rate. The initial model included all the explanatory variables described above. Subjects and the items were taken as random effects. From the group of extralinguistic factors, OCCUPATION reached a marginal significance ($\chi2 = 3.1101$, $df = 1$, $p = 0.077$). AGE has also reached a significance level ($\chi2 = 4.9821$, $df = 1$, $p = 0.025$). PRONLENGTH has reached, as expected, the significance level ($\chi2 = 5.9334$, $df = 1$, $p = 0.014$), and this effect is stable both with the raw pronunciation length and the speech rate measured in phon/sec. GENDER did not reach the significance threshold ($\chi2 = 0.0579$, $df = 1$, $p = 0.8$).

In the lexical factors group, four different frequency measures were considered: the FREQUENCY, the FREQSCHWAWORD and FREQSECMORPH, RELFREQBASE. Unexpectedly, none of these four frequency measures showed a significant effect: FREQUENCY ($\chi2 = 0.2808$, $df = 1$, $p = 0.596$), FREQSCHWAWORD ($\chi2 = 0.2527$, $df = 1$, $p = 0.615$), RELFREQBASE ($\chi2 = 0.1084$, $df = 1$, $p = 0.742$), and FREQSECMORPH ($\chi2 = 0.6175$, $df = 1$, $p = 0.432$). Moreover, FAMILIARITY also reached just a marginal significance ($p = 0.098$). ITEM TYPE has reached a significance level ($\chi2 = 7.5685$, $df = 1$, $p = 0.0059$). REPETITION reached a marginal significance in the model, but only the difference between the first and the second repetition ($p = 0.0602$). However, removing REPETITION from the model did not significantly impact the model fit ($\chi2 = 3.7571$, $df = 2$, $p = 0.1528$).

From the last group of variables, the PRECEDINGSOUND positively affects the schwa deletion rate ($\chi2 = 22.287$, $df = 3$, $p = 5.686e\text{-}05$). POSITION ($\chi2 = 0.5576$, $df = 1$, $p = 0.4552$) did not render any significant effect.

Interaction between FAMILIARITY*ITEM TYPE reaches a marginal significance ($\chi2$ = 3.277, *df* = 1, *p* = 0.07). Moreover, the interaction between OCCUPATION*AGE rendered a significant effect ($\chi2$ = 3.9495, *df* = 1, *p* = 0.046). The results of the final model are illustrated in Table 14 below.

Table 14. Results of the pre-nasal schwa elision mixed-effects regression modelling in experiment II

	Estimate	**Std Error**	**z Value**	***p***
Intercept	2.43899	0.6763	3.606	0.0003
Age	0.0834	0.1388	0.601	0.5477
Occupation:Non	-0.47008	0.2339	-2.010	0.0444
Familiarity	0.5856	0.2624	2.231	0.0256
Item Type:Term	3.0266	0.9523	3.178	0.0014
PronLength	-0.1619	0.0692	-2.343	0.0191
Repetition2	0.2770	0.1466	1.888	0.0589
Preceding sound:Plosive	-5.0500	0.8691	-5.811	6.22e-09
Preceding sound:Nasal	-7.6313	1.3977	-5.460	4.77e-08
Preceding sound:Taps	-5.4730	1.2136	-4.510	6.49e-06
Age*Occupation:Non	0.4349	0.2121	2.050	0.04032
Familiairity*Item Type:Term	0.3073	0.1702	1.806	0.07095

This data shows that AGE and FAMILIARITY have a positive correlation with the schwa elision rate. PRONLENGTH (which can be treated as a measurement for speech rate) has a negative correlation with the schwa elision rate. The results are in line with the expected findings. The members of the law group were more likely to delete the schwa word. The fricatives created the most favourable phonological context for the schwa elision from the four possible sounds that preceded the schwa.

8.2.1.2 Analysis of /t/ elision in the selected words

Studies focusing on word-final /t/ and /d/ deletion in English were carried out by Bybee (2002b) and Jurafsky et al. (2001a). Kohler is a prominent scholar investigating phonetic reduction in German. One of the most typical reduction processes in German is /t/ elision. There are a limited number of words in which the elision occurs, and as such, only 10 words were eligible for analysis, 6 words from the general group and 4 from the field of law. Every participant pronounced the words three times, which resulted in 960 recordings containing the /t/ sound in word-internal position. Whilst 10 words do not provide sufficient data from which to draw straightforward conclusions, the data allows us to observe initial trends.

All the 960 recordings included the /t/ in the middle of a consonantal cluster. These clusters were the following: [fts], [ftl], [çts], [ltʃ] and [stm]. Not all of the consonantal clusters listed are found in Kohler (1977). According to Kohler, one of the most favourable environments for the /t/ elision is if the /t/ is preceded by an /n/ or /l/ and followed by an /s/. None of the words from the word naming task had this phonological environment. One of the words, *Staatsanwaltschaft* was the closest to this environment with the only modification that instead of the apico-alveolar /s/, the /t/ was followed by a palato-alveolar /ʃ/. The combination of the apico-alveolar /t/ and of the palato-alveolar /ʃ/ results in the palato-alveolar affricate /tʃ/. Cases, where the apico-alveolar merged with the palato-alveolar sound, were not regarded as cases of the /t/ elision. A more accurate way of describing this process is coarticulation. Out of the 10 selected words, two contained the /ftl/ consonantal cluster, but all the participants realised the /t/ in these clusters. Besides the clusters that are enlisted by Kohler (1977), two additional clusters were analysed /çts/ and /stm/ in the following words *Gerichtsverfassungsgesets, Grundrechtseingriffe, Rechtsmittelverfahre,* and *Verdiestmöglichkeit*. After excluding the five general words which did not show any tendencies toward the elision of the apico-alveolar /t/, five words remained, which resulted in 480 recordings.

The predictors of the /t/ elision are the same as those used for the analysis of the schwa elision. The only difference is that only the frequency of the whole word was taken into account. As per the previous segmental context, the /t/ was preceded by the /t/ was preceded by a palato-dorsal /ç/, a palato-alveolar /l/ and an alveolar /s/.

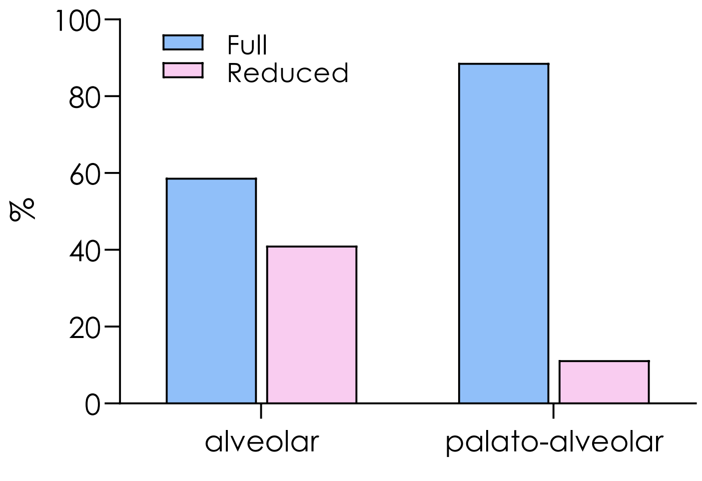


Figure 27. The effect of the previous sound on the rate of /t/ elision

The following sounds included the apico-alveolar /s/, the palato-alveolar /ʃ/ and the bilabial /m/. The sounds were coded as categorical variables describing the places of articulation. Additionally, a category describing whether the preceding and following sounds were homorganic or non-homorganic with the /t/ was added to the analysis. This was respectively coded as a categorical value.

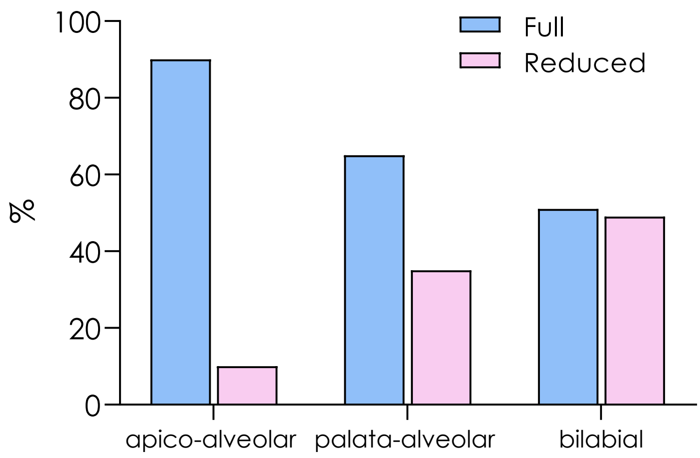


Figure 28. The effect of the following sound on the rate /t/ elision

It should be highlighted that this analysis does not make definitive conclusions regarding the interconsonantal /t/ elision; it only analyses the observed tendencies. In the first step of the analysis, all recordings were manually examined in order to determine the presence or absence of the /t/ sound. The analysis consisted of two steps: initially, the spectrogram representing the pronunciation of the word was examined, and then the TextGrid file (the output file of the WebMaus Basic) containing information on the length of the pronunciation of each sound was compared with the results obtained from the analysis of the spectrogram. If, on the spectrogram, the /t/ sound was neither visible nor audible, and the software did not assign any temporal characteristic to the /t/ sound in the TextGrid file, then and only then was the word considered to illustrate the /t/ elision. This procedure was carried out with all the word occurrences. In the end, it was established what the percentage of /t/ elision in the selected words was. The results are summarised in Table 15 below.

Table 15. Percentage of interconsonantal /t/ elision in the selected words

Word	***Percentage of /t/ elision***
Gerichtsverfassungsgesetz	20.8 %
Grundrechtseingriffe	0.04 %
Rechtsmittelverfahren	0.04 %
Staatsanwaltschaft	35 %
Verdienstmöglichkeit	49 %

8.2.1.2.1 Modelling phonetic reduction

The analyses were performed using the generalised linear model *glmer()* from the package *lme4*. As the first step, all the categorical values were transformed into factors. Continuous variables were first scaled using the transform function. The selection of the explanatory variables was made using the backward method. Participants and words were regarded as random effects. The initial global modal has revealed a high correlation between FREQUENCY and FOLLOWINGSOUND. The reason for the high correlation is that the most favourable environment for reduction was when a bilabial sound followed the/t/, and these words had the highest frequency.

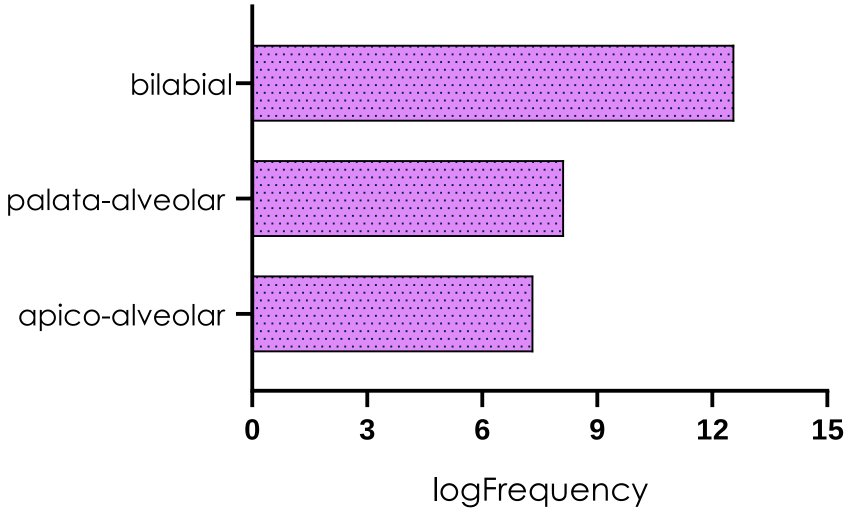


Figure 29. The correlation between frequency and the following sound

For this reason, the two explanatory variables could be used in the same model. Two additional global models were fit to the data with the only difference that the first one included PREVIOUS SOUND and FREQUENCY and the other FOLLOWING SOUND without FREQUENCY. The two models have equal predictive powers (R^2 = 30.9), and AIC values are also almost identical (model 1 = 421.09, model 2 = 420.83).

The model selection has shown that OCCUPATION, surprisingly, rendered no significant effect on the word-internal /t/ reduction (p = 0.8529). From the extra-linguistic variables, AGE ($\chi2$ = 7.1613, df = 1, p = 0.007) and PRONLENGTH ($\chi2$ = 2.9336, df = 0, $p < 0.001$) reached the significance level as a predictor of the rate of the /t/ elision.

The group of lexical variables included: FREQUENCY, FAMILIARITY, REPETITION, and ITEM TYPE. ITEM TYPE cannot be regarded as a reliable explanatory variable because the distribution among the legal terms and distractors is not equal. Even though there is only one word belonging to the category of general words, this word was the most frequently reduced. This emerges from Table 16. Unexpectedly, REPETITION turned out to be an insignificant explanatory variable ($\chi 2 = 0.0167$, df = 1, p = 0.89). The last lexical variable, FREQUENCY, rendered a significant effect on the rate of /t/ deletion as expected ($\chi 2 = 21.419$, df = 4, $p < 0.001$). FAMILIARITY was on the verge of significance ($\chi 2 = 2.9557$, df = 1, p = 0.08). The results of the final model are demonstrated in Table 16 below.

Table 16. Results of the interconsonantal /t/ elision mixed-effects regression modelling in experiment II

Predictor	**Estimate**	***t* Value**	***p***
Intercept	5.0240	3.345	<.001
Age	0.4773	2.873	< .01
Occupation:Law	0.5231	1.356	.17
Familiarity	0.3114	1.668	.085
PronLength	-0.4034	1.719	.043
Frequency	16.1060	4.887	<.001
Preceding sound (palata-alveolar)	-38.4498	-4.730	<.001
Preceding sound (apico-alveolar)	2.3067	6.707	<.001

8.2.2 Experiment IV

8.2.2.1 Reduction of the prenasal schwa

For the analysis of the prenasal schwa reduction, 27 words were selected that had the phonetic environment facilitating the reduction processes. Out of the 27 words, 18 were adjectives, and the remaining were nouns. This resulted in 1228 recordings. The process of determining the presence of the schwa sound was identical to that described in experiment II.

The group of linguistic factors included: frequency, bigram frequency, familiarity, and place of articulation of previous sound - PLACESOUND. The manner of articulation included the following sounds: MANNERSOUND, TRIGGER (a categorical variable that induces the syllable in which the schwa sound occurs: *-ten, -len, -chen, -del, -gen, -den, -schen*), TP_FORWARD, TP_BACKWARD, RECALL (whether the word was recalled or not during the sentence recall). The social and cognitive factors were identical to those described in modelling temporal reduction.

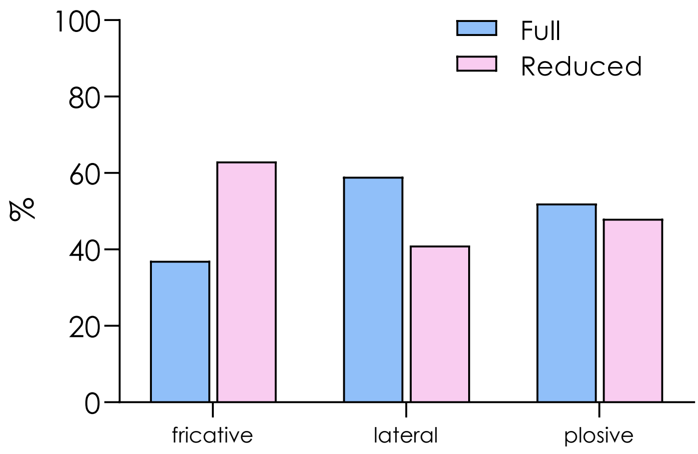


Figure 30. The effect of the manner of articulation on the rate of schwa reduction in Experiment IV

Figure 30 reveals that fricatives were the most often reduced, followed by plosives. There is no significant difference between plosives and laterals.

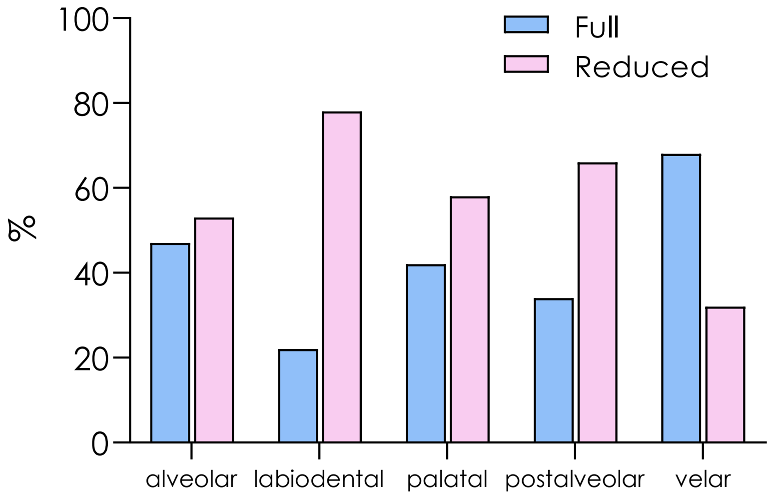


Figure 31. The effect of place of articulation on the rate of schwa reduction in Experiment IV

Figure 31 illustrates that there are no huge differences between the different places of articulation. The labiodental sounds have the highest rate of reduction, and the velar sounds have the lowest. Alveolar, palatal, and postalveolar sounds showed a similar pattern of reduction.

8.2.2.1.1 Modelling of phonetic reduction

Modelling was done using the *glmer()* function. Frequency-connected variables were log-transformed, FAMILIARITY, VERBAL and NONVERBAL PROCESSING were scaled. The initial global model included all the social, cognitive, and linguistic variables described above. To ensure that the correlation between the explanatory variables does not influence the model fit, their VIF values were checked, and all of them were below the threshold level of 5. The backward selection method was used as established in the previous analysis.

From the social factors, neither OCCUPATION (χ2 = 0.2131, df = 1, p = 0.6443), nor AGE (χ2 = 1.4059, df = 1, p = 0.2357) reached significance to predict phonetic reduction. Out of the cognitive factors, NONVERBALPROCESSING (χ2 = 0.3621, df = 1, p = 0.948) did not reach the significance level, and VERBALPROCESSING (χ2 = 20.765, df = 1, p = 5.191e-06) turned out to be a significant predictor of schwa reduction. PRONLENGTH (χ2 = 1.0349, df = 1, p = 0.309) did not reach significance, contrary to expectation.

The analysis of linguistic predictors reveals some unexpected results. Only 3 of the variables reached the significance level: RECALL (χ2 = 216.53, df = 1, p = 2.2e-16), FAMILIARITY (χ2 = 50.441, df = 1, p = 1.228e-12), TP_FORWARD (χ2 = 13.199, df = 1, p = 0.0002). None of the other predictors rendered a significant effect of schwa reduction: ITEM TYPE (χ2 = 0.0529, df = 1, p = 0.8181), FREQCOLLOC (χ2 = 0.0529, df = 1, p = 0.6607), TP_BACKWARD (χ2 = 0.0531, df = 1, p = 0.8178), FREQUENCY (χ2 = 0.9349, df = 1, p = 0.336), MANNERSOUND (χ2 = 1.9446, df = 1, p = 0.1632), PLACESOUND (χ2 = 10.279, df = 8, p = 0.246). The insignificance of the last two predictors, MANNERSOUND and PLACESOUND, was expected, based upon the visualisation of the distribution of the rate of reduction of the manner and place of articulation. The four significant predictors alone can explain 43% of the variance. The final model is summarised in Table 17.

Table 17. Summary of the final model with schwa reduction as the dependent variable

	Estimate	**Std Error**	***t* Value**	***p***
Intercept	-0.95650	0.22638	-4.225	2.39e-05
VerbalProcc	-0.61044	0.11902	-5.129	2.92e-07
Recall:Yes	2.68305	0.19428	13.810	< 2e-16
TPForward	0.66987	0.17212	3.892	9.95e-05
Familiarity	0.66970	0.09541	7.019	2.24e-12

8.3 Strength of memory

8.3.1 Experiment III

In the following experiment, the analysis of the verbatim recall will be discussed. The dependent variable is the accuracy of recall of the target adjective-noun pairs. It is coded as a binary categorical variable. The same explanatory variables were used as those to analyse temporal reduction, with one additional variable. The additional variable describes the strength of associations between modifiers and nouns that language users have. It was elicited from an online survey completed by 100 participants. The methodology was precisely the same that was used to test the high-transitional probability bigrams. After reading the given adjective, participants were asked to write down the first noun that came to their minds. The adjectives showed gender marking. The rationale behind carrying out an additional study was to compare the corpus-based transitional probabilities with actual associations that language users have. ASSOCIATIONS is a continuous variable that indicates the percentage of participants who provided the expected noun.

8.3.1.1.1 Modelling the strength of memory representation

Model fitting comprised the following steps: first, a global modal was fitted to the data with ACCURACY as a dependent variable and SUBJECT and ITEM as random effects. Using the *dregdge()* function, an automatic model selection was first implemented and then compared to backward stepwise variable selection for both random and fixed effects. All the continuous variables were rescaled.

First, the explanatory variables that did not reach the significance level in the global model were removed. OCCUPATION ($\chi2 = 0.1596$, df = 1, p = 0.3492), GENDER ($\chi2 = 0.2488$, df = 1, p = 0.6179) from the group of linguistic variables did not render a significant effect on ACCURACY. AGE ($\chi2 = 7.7787$, df = 1, p = 0.05), CRIME SERIES ($\chi2 = 4.1151$, df = 1, p = 0.0425) reached the threshold to be significant predictors.

From the group of linguistic variables, none of the frequency measures, including FREQNOUN ($\chi2 = 0.2368$, df = 1, p = 0.6266), FREQCOLLOC ($\chi2 = 1.415$, df = 1, p = 0.2342), FREQMODIFIER ($\chi2 = 0.9473$, df = 1, p = 0.6266), DISPERSION ($\chi2 = 2.1938$, df = 1, p = 0.3304), rendered a significant effect on the accuracy of recall. Unexpectedly, none of the association measures reached the significance threshold: TP ($\chi2 = 0.5523$, df = 1, p = 0.4574), LOG-LIKELIHOOD ($\chi2 = 0.3743$, df = 1, p = 0.5407), MUTUAL INFORMATION ($\chi2 = 0.3944$, df = 1, p = 0.53). The three variables that passed the Bonferroni-corrected significance level of 0.005 from the group of linguistic variables are ITEM TYPE ($\chi2$ = 12.988, df = 1, p = 0.0003), CONDITION ($\chi2$ = 35.477, df = 2, p < 0.0001), FAMILIARITY ($\chi2$ = 11.359, df = 5, p = 0.047).

The group of cognitive factors produced some unexpected results. The two memory tests that were expected to predict the ACCURACY of recall, namely NUMBERSPAN ($\chi2$ = 0.7321, df = 1, p = 0.3922), WORDSPAN ($\chi2$ = 1.4364, df = 2, p = 0.4876), did not render any significant effects. Additionally, neither ASSOCIATIONS ($\chi2$ = 0.1349, df = 1, p = 0.7134), nor READING ACCURACY ($\chi2$ = 2.33, df = 1, p = 0.1267), nor MENTPROCACC ($\chi2$ = 2.241, df = 1, p = 0.1344) reached the threshold as a significant predictor of ACCURACY. The only variables from the group of cognitive factors that render a significant effect on the performance of the participants are MENTPROCRT ($\chi2$ = 14.746, df = 1, p < 0.0001), and READING ($\chi2$ = 6.963, df = 1, p = 0.008).

The final model was extended with an interaction between ITEM TYPE and OCCUPATION, but it did not render any significant effect ($\chi2$ = 0.8276, df = 2, p = 0.6611). Additionally, two random slopes were added to the final model with a by-SUBJECT random slope on ITEM TYPE, and by-ITEM random slope on FAMILIARITY. However, the random slopes did not improve the model (AIC model without random slopes = 1324.2, with random slopes = 1361.0). The results of the final model are shown in Table 18.

Table 18. Summary of the mixed-effects regression model fitted to the accuracy data

Predictor	**Estimate**	**z Value**	***p***
Intercept	2.9970	2.872	0.00408
Age	1.3106	3.199	0.00138
Crime series	0.9605	2.358	0.01839
Item type (terminology)	-1.3360	-5.539	3.05e-08
Condition (high)	1.1687	5.979	2.25e-09
Condition (frequency)	0.6235	2.448	0.01437
Familiarity (score 7)	1.1450	2.211	0.002704
Mental Processing RT	-4.4158	-4.986	6.17e-07
Reading	0.9770	2.768	0.00563

The result of the modelling reveals that older participants were more likely to recall the selected bigrams correctly. 75% of the participants were younger than 25 years of age. However, the remaining 25% of the participants still performed better than the younger participants. Those participants who indicated that they spent some time watching crime series were better at remembering the bigrams. There is a clear difference between the general and legal bigrams; participants preferred the general bigrams over the legal terminology. As expected, bigrams from the high-transitional probability condition were recalled the most accurately, followed by the high-frequency bigrams. Familiarity with the items enhanced recall, and slower reaction times in the mental processing speed task facilitated recall. Reading has shown a reversed direction as expected; the more time participants spend reading the given texts, the more accurately they have recalled the bigrams.

8.3.2 Experiment IV

Recordings from the sentence recall task were analysed. The recall accuracy was coded as a categorical variable that included *recalled, not recalled, replaced*, and *partially recalled* adjective-noun pairs. The social and linguistic explanatory variables are identical to those used to model temporal reduction in experiment IV. The cognitive variables included IMPLICIT and EXPLICIT LEARNING and NON-VERBAL PROCESSING.
IMPLICIT LEARNING is a continuous variable that shows the percentage of correct answers that the participants have achieved in the embedded triplets task. EXPLICIT LEARNING is also a continuous variable showing how many pairs the participants remembered during the paired association learning task. NON-VERBAL PROCESSING is the average time participants needed to process the non-verbal information during the cued task-switching experiment. It is a continuous variable that is log-transformed for the analysis. A further cognitive factor that was extracted from the cued task-switching task was the task-switching cost measured in milliseconds. The variable is referred to as TASK SWITCHING and is coded as a continuous variable.

8.3.2.1 Modelling the strength of memory representation

Generalised mixed-effects regression models can only accept a binary categorical variable as the dependent variable. For this reason, the four categories had to be reclassified into two; everything that was coded as partially recalled, replaced, and not recalled was coded as not recalled.
First, a global model was fitted to data that included all the explanatory variables. The variables with the highest p-values were removed using the backwards selection method, and the resulting models were compared. The manual and automatic model selection resulted in the same explanatory variables.

From the linguistic factors, FREQCOLLOC ($\chi2$ = 0.1364, df = 1, p = 0.7119), NCD ($\chi2$ = 2.2068, df = 1, p = 0.1374), TP_FORWARD ($\chi2$ = 0.007 df = 1, p = 0.933) did not render any significant effect on determining the strength of memory representation. The linguistic predictor that reached significance level included FREQNOUN ($\chi2$ = 8.3573, df = 1, p = 0.003841), TP_BACKWARD ($\chi2$ = 4.1782, df = 1, p = 0.04095), FAMILIARITY ($\chi2$ = 13.137, df = 1, p = 0.0002896), WORDLENGTH ($\chi2$ = 29.444, df = 2, p = 4.04e-07), ITEM TYPE ($\chi2$ = 6.2596, df = 1, p = 0.012). Out of the social factors, OCCUPATION ($\chi2$ = 47.336, df = 1, p = 5.98e-12), GENDER ($\chi2$ = 11.37, df = 1, p = 0.0007463) seem to influence the strength of memory representation. Regarding the cognitive factors, NON-VERBAL PROCESSING ($\chi2$ = 0.5306, df = 1, p = 0.4664) and TASK SWITCHING ($\chi2$ = 0.1655, df = 1, p = 0.6841) did not reach significance. All the other explanatory variables show a significant effect on the strength of memory representation: VERBAL PROCESSING ($\chi2$ = 8.9207, df = 1, p = 0.00282), EXPLICIT LEARNING ($\chi2$ = 12.722, df = 1, p = 0.0003613), IMPLICIT LEARNING ($\chi2$ = 54.859, df = 1, p = 1.295e-13), SPEECH RATE ($\chi2$ = 199.98, df = 1, p = 2.2e-16).

As the next step, the interactions between the different variables were explored. The interaction between GENDER and OCCUPATION did not render a significant effect ($\chi2$ = 1.4483, df = 1, p = 0.2288). SPEECH RATE and FAMILIARITY, as expected, showed a highly significant effect.

The three-way interaction among SPEECH RATE, FAMILIARITY and FREQNOUN did not show a significant effect. Surprisingly, the three-way interaction between SPEECH RATE, FAMILIARITY and WORD LENGTH rendered a significant effect on the model fit. The final model with all the significant explanatory variables and interactions is summarised in Table 19 below.

Table 19. Summary of the final model with Recall as a dependent variable

Predictor	Estimate	Std Error	z Value	*p*
Intercept	0.676811	0.5536	1.222	0.2215
Occupation:Non	-0.440481	0.0645	-6.820	9.08e-12
Gender:Male	0.220921	0.0659	3.352	0.000803
Verbal Processing	0.087686	0.0310	2.821	0.004782
Explicit Learning	0.104117	0.0305	3.411	0.000648
Implicit Learning	0.224406	0.0308	7.269	3.62e-13
Speech Rate	0.054281	0.0481	1.127	0.259886
Familiarity	0.156248	0.3926	0.398	0.690717
Word Length	-0.382292	0.0697	-5.479	4.29e-08
FeqNoun	0.188450	0.1006	1.873	0.06109
TP_Backward	0.241285	0.1028	2.345	0.019013
ItemType:Terminology	-0.55855	0.21953	-2.544	0.010951
SpeechRate*Familiarity	-0.006779	0.0402	-0.168	0.86638

SpeechRate*WordLength	0.023579	0.0059	3.951	7.78e-05
Familiarity* WordLength	-0.114152	0.0520	-2.192	0.028388
SpeechRate*Familiarity*LengthNoun	0.011772	0.0051	2.300	0.021459

A multinomial regression was fitted to the data using the method outlined by Levshina (2015), using the function polytomous() to evaluate how the different predictors influenced the recall of the items. This function enables all four levels to be taken into account all the four levels: *recalled, not recalled, partially recalled* and *replaced*. "Partially recalled" refers to those adjective noun-pairs where only either the adjective or the noun was recalled. Those adjective-noun combinations where either the adjective or the noun was replaced by another word were coded as replaced. The explanatory values from the final model were fitted to the model. The summary of the multinomial regression is presented in Table 20 below.

Table 20. Summary of the multinomial regression with estimates and p-values

predictor	not recalled	partially recalled	recalled	replaced
Group:Non	0.4951	0.2439	-0.373	(-0.03672)
p-value	0.000	0.011	0.000	0.517
Gender:Male	(-0.1216)	(-0.05315)	0.2082	-0.1683
p-value	0.110	0.590	0.000	0.013
Verbal Processing	(-0.002238)	(0.0009081)	0.002842	(-0.002719)
p-value	0.065	0.653	0.003	0.033
Explicit Learning	-0.02561	(-0.003186)	(0.008102)	0.01578
p-value	0.001	0.766	0.176	0.503
Implicit Learning	-0.01636	-0.008634	0.01191	(0.002926)
p-value	0.000	0.004	0.000	0.163
Familiarity	-0.1174	(-0.02956)	0.09884	(-0.01027)
p-value	0.000	0.487	0.000	0.812
WordLength	0.05943	0.07711	-0.06325	(-0.01197)
p-value	0.000	0.000	0.000	0.292
FeqNoun	0.05943	0.07711	-0.06325	(-0.01197)
p-value	0.003	0.989	0.003	0.409
TP_Backward	-8.246	-4.907	3.335	3.825
p-value	0.000	0.000	0.001	0.001
ItemType:Term	0.3027	0.2154	-0.6001	0.4697
p-value	0.000	0.034	0.000	0.000

9 Findings and Discussion

9.1 Finding 1

One of the main focuses of the four experiments conducted within the framework of this dissertation was on understanding whether or not differences in the experience with a particular register contribute to different judgements about the familiarity with the selected items characteristic of that register. Due to the participants' different experiences in the domain of law, the two groups were expected to differ in their usage of words and collocations customary to that domain. According to the usage-based approach, this difference in the amount of experience leads to different mental representations. The hypothesis that emerged from this assumption was that the strength of the linguistic representations should be reflected in the participants' familiarity judgements. Moreover, the familiarity scores were expected to determine the participants' performance in the different tasks.

In order to validate this hypothesis, in 3 out of the 4 experiments, participants were asked to rate how familiar they were with the stimuli. Subsequently, the relationship between their metalinguistic judgements and temporal and phonetic reduction and strength of memory representation was investigated. This investigation indicates that familiarity judgements are more powerful in explaining the observed variation in the participants' performance than corpus-based measures such as frequency, transitional probability, or contextual diversity.

The analysis of the experiments that used familiarity as a predictor of the participants' performance led to the following conclusions:

9.1.1 Familiarity scores are valid measures of the participants' previous language exposure and usage intensity.

These scores are valuable metalinguistic data and can be viewed as language-user-dependent frequency measures. Corpus-based frequencies provide valuable information about the general tendencies observed in language. On the other hand, familiarity scores can help explain the variation observed in language comprehension and production. The two measures are unquestionably related as those words that occur more often in the language will be assigned higher familiarity scores. Moreover, familiarity scores are even more valuable when those registers investigated which are poorly represented in a general corpus.

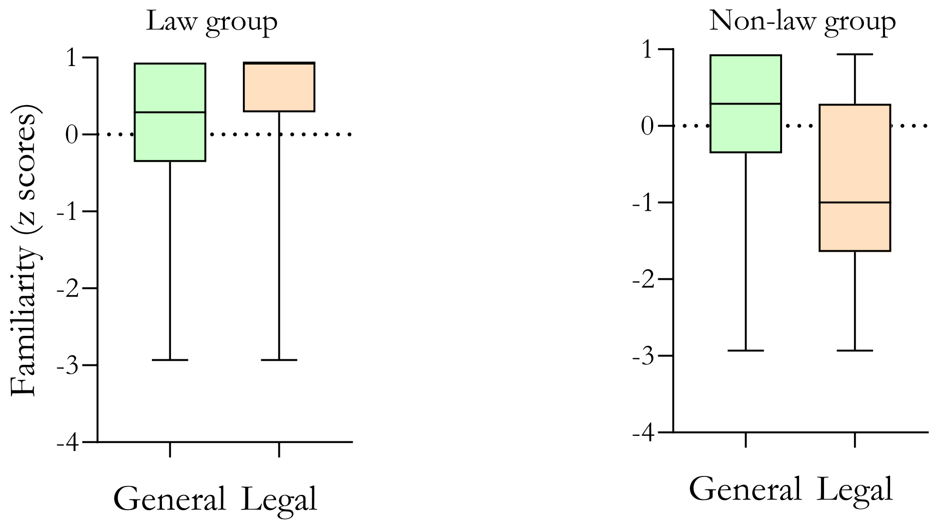


Figure 32. Average familiarity scores (converted to Z scores) by group and item type. The figure is generated based on the familiarity scores obtained during experiment II

Figure 32 supports the finding and shows that the experience lawyers have with legal words is reflected in the higher scores assigned to the legal words. Furthermore, figure 32 reveals that the law group members gave lower scores to the general words than the law words. This result can be explained by the fact that legal words have more robust representations in the participants' mental lexicon. The legal stimuli used for the experiment were extracted from the lecture corpus. It was expected that all the participants from the law group frequently used or heard/saw these words. Another possible explanation is the low frequency of the general words, as they were selected to match the length of the legal words. Following Zipf's law, there is a reversed correlation between word length and frequency. Most legal words were polymorphemic compounds; thus, the equally long general words came from the low-frequency group. The non-law group's familiarity scores further support the finding; they rated the legal words as less familiar than the general words, reflecting their limited experience with the legal words.

Familiarity scores from experiment IV follow the tendency highlighted above; however, the difference between the ratings of the law group on the legal and general words is less distinct than in the case of experiment II. The results are visualised in Figure 33.

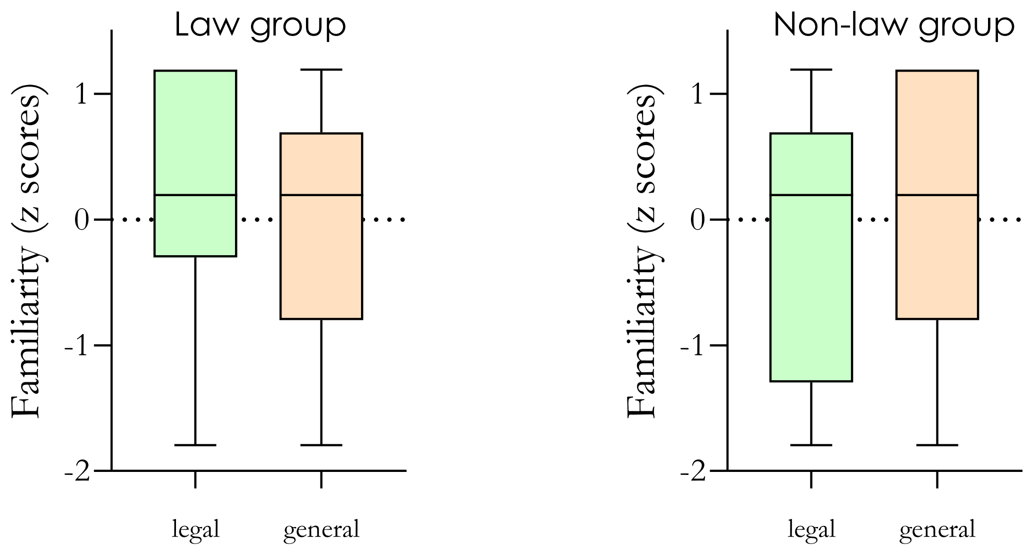


Figure 33. Average familiarity scores (converted to z scores) by group and item type. The figure is generated based on the familiarity scores obtained during experiment IV

It should be mentioned that participants of experiment IV were law students who had only limited experience with legal collocations compared to employees of a law chair. The general collocations were selected to demonstrate high-frequency, and transitional probability and legal collocations were less specific than in the case of experiment II. Thus, despite the surface differences between the familiarity scores obtained in experiments II and IV, they show the expected tendency in both cases, i.e., familiarity scores mirror the participants' experience with the selected items. A further difference between the two experiments is that, on average, participants gave higher familiarity scores. The higher frequency of the selected collocation can account for these higher scores.

9.1.2 Familiarity was the only significant explanatory variable in all the statistical models across the different paradigms.

Occupation as a predictor of the participants' experience in most experiments also significantly predicted the participants' performance on the different tasks. However, this effect of occupation is more significant in experiment II than in IV, as shown in figure 34. The other linguistic predictors did not show the same consistent impact on temporal, phonetic reduction, and strength of memory representation as familiarity scores did.

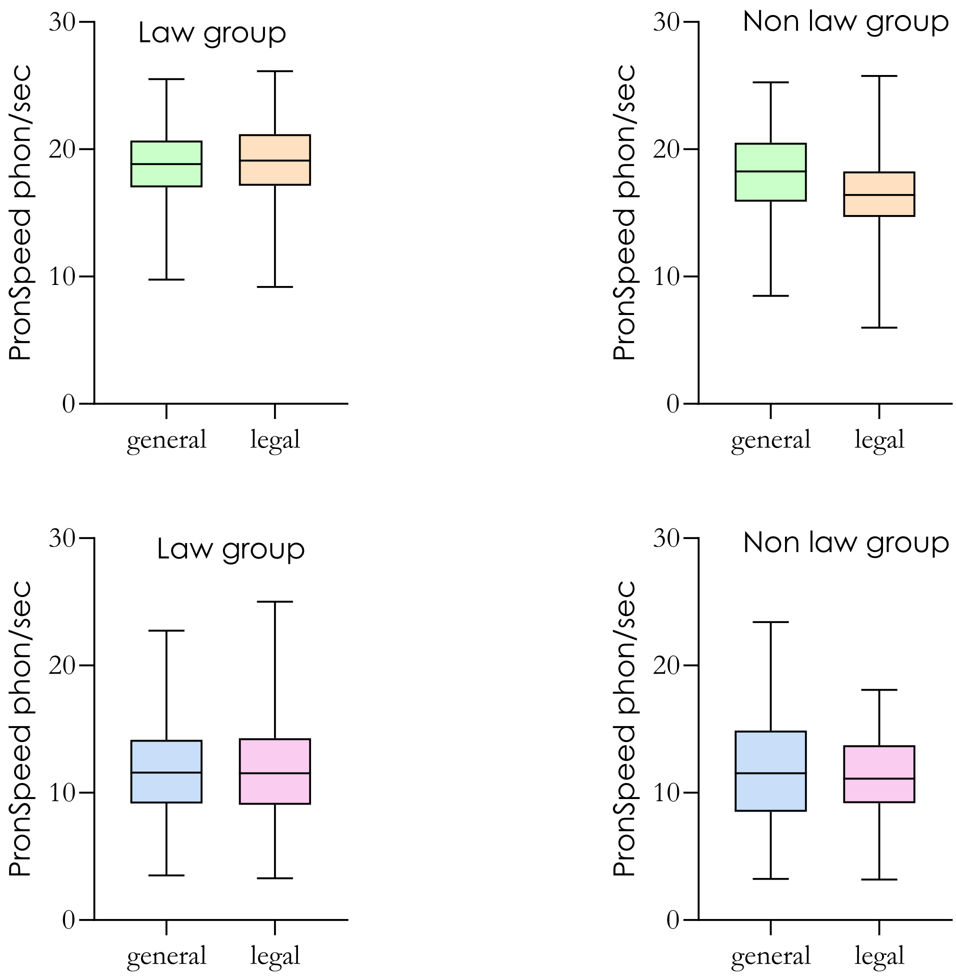


Figure 34. The effect of occupation on the pronunciation speed across groups and item types. The upper two graphs stem from experiment II, the lower two from experiment IV

Figure 34 illustrates that the law group pronounced the legal words slightly faster than the general words in experiment II. This difference is even more considerable in the case of the non-law group, which further supports the above-highlighted hypothesis about the connection between the experience with the language and the strength of memory representation. It is interesting to note that participants had significantly higher speech rates in experiment II than in IV. There are two possible explanations for this observation. First of all, in experiment II, participants were exposed to single words and instructed to pronounce them as fast as possible. Second, participants in experiment II, as has been already pointed out, had more experience with the legal word than those in experiment IV. Furthermore, figure 34 shows that there is no significant difference between the pronunciation rate across the groups and items in experiment IV. The inspection of the correlation between the familiarity scores and temporal reduction revealed that familiarity scores are more significant predictors of the participants' performance than group membership.

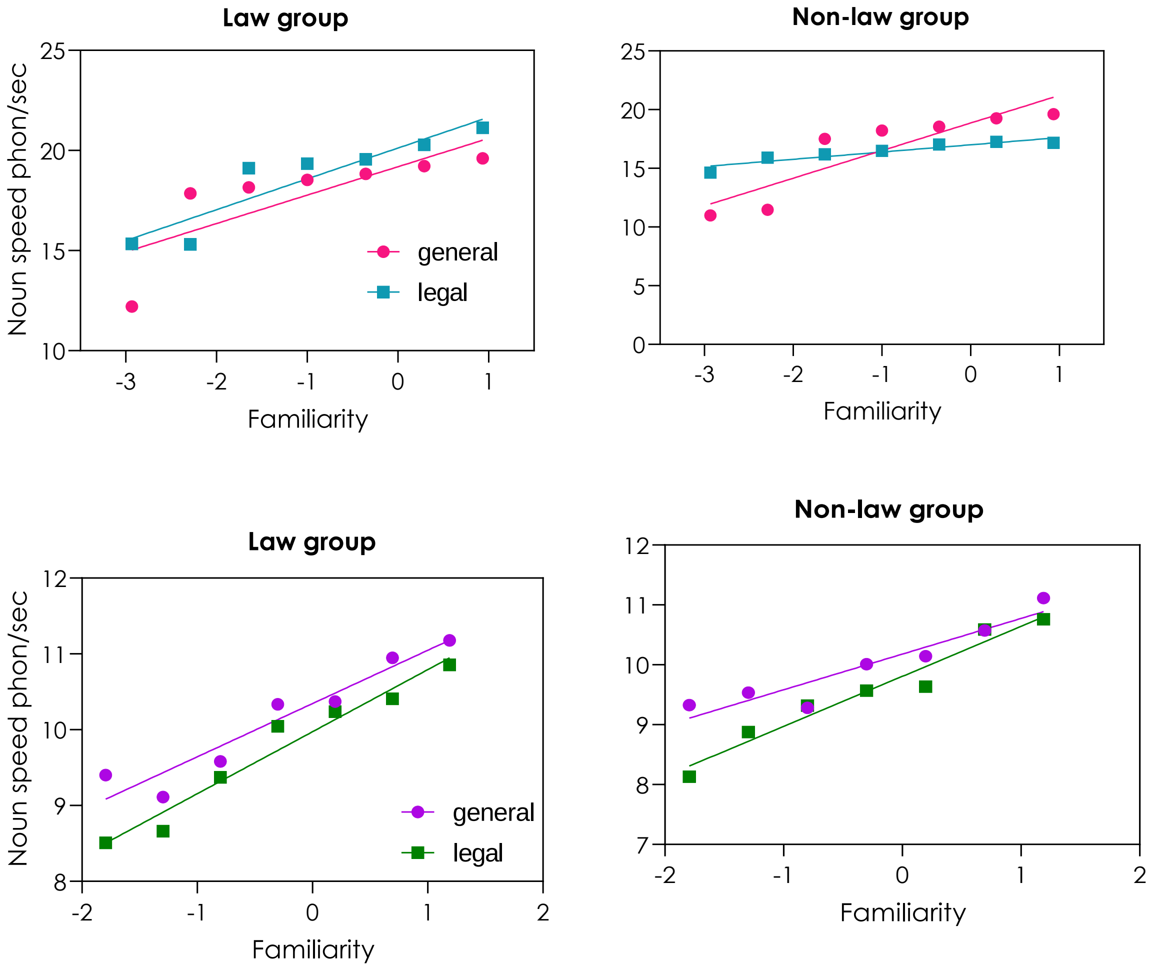


Figure 35. Correlation between speech rate and familiarity ratings. The upper two figures are based on the data from experiment II and the lower two from experiment IV. The separate dots are based on the average speech rate for each familiarity rating

Figure 35 reveals that, as predicted, there is a positive correlation between the participants' speech rate and their familiarity scores. This correlation is more linear in the case of the law group, as the two left-side figures demonstrate. Moreover, the upper right-side scatterplot illustrates that it is more difficult to judge the familiarity of the low-frequency items. In experiment II, members of the non-law group could not objectively judge how often they had seen/heard and used the legal items. The regression line is almost flat, and there is no significant difference in the pronunciation speed between words that were assigned a high or a low familiarity score. The tendency is straightforward in all the other cases; higher familiarity scores are reflected in faster speech rates. The lower graphs also support this assumption about the difficulty of judging less familiar items since the stimuli for experiment IV was selected to be frequent or have a high-transitional probability. In these cases, the regression line is more exponential.

The analysis of the prenasal schwa also revealed that words with higher familiarity scores were more often reduced. As was expected, members of the law group produced more reduced legal words, and members of the non-law group reduced the general words more often.

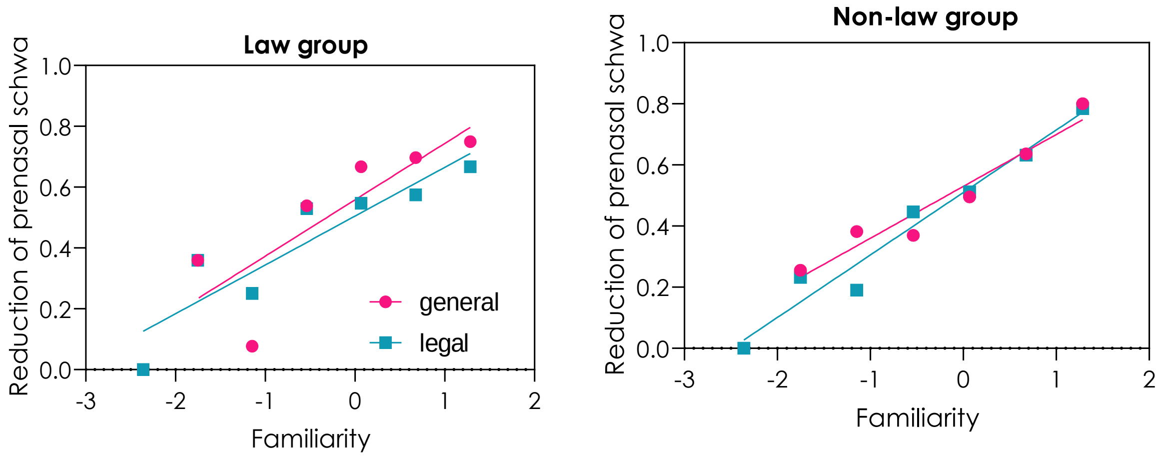


Figure 36. The effect of familiarity on the rate of prenasal schwa reduction

One unanticipated finding was that the correlation between group membership and the speech rate in experiments II and IV does not hold for the memory recall task; thus, members of the law group did not recall the legal collocations more accurately than the general collocations. It was hypothesised that previous experience with the words measured as group membership would influence the participants' performance on all the tasks. This suggests that the assumption that someone's experience with language can be determined based on their social background is misleading. As figures 35 and 36 show, familiarity scores provide more reliable information on that matter.

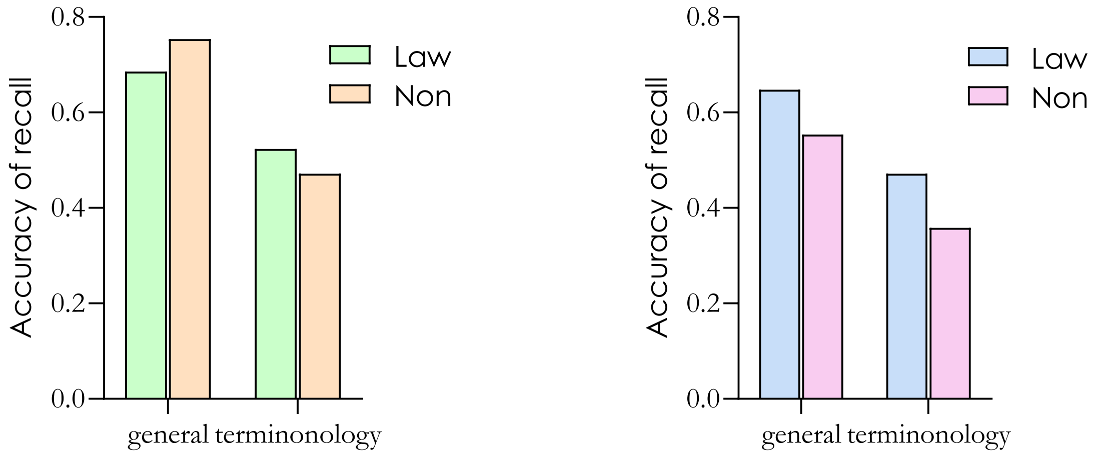


Figure 37. Social background and language experience. The left figure represents experiment III, and the right-side figure experiment IV

Figure 37 above shows that law group members had more robust memory representations for the general words since they recalled more general than legal collocations. Members of the law group in both experiments outperformed the other group. On average, they recalled significantly more legal and general collocations. The reason why law students, on average, recalled more items could be due to the fact that they have to learn many things by heart during their studies. Thus, their memory is trained to store and recall items after a short exposure.

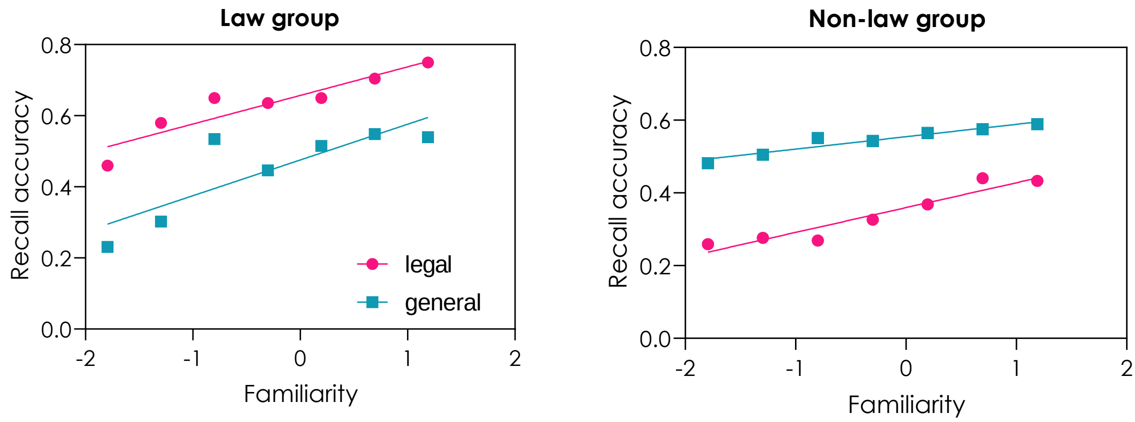


Figure 38. Correlation between familiarity scores and accuracy of recall. The figures are based on Experiment IV

From the data in Figure 38, it can be concluded that familiarity scores can more accurately determine prior language experience than the social background of the participants, i.e., their group membership. Moreover, the result suggests that familiarity scores can account for the variation observed in homogeneous groups.

Four important implications follow from these findings. First, they indicate that familiarity judgements are valuable psycholinguistic data that let us researchers gain a glimpse into the mental representations language users have, and these scores are also likely to reflect the strength of these representations. Second, the findings align with the main postulations of the usage-based theories according to which language usage is the main driving force of language acquisition in general and entrenchment in particular. Third, language users can better judge the familiarity of frequent linguistic items than low-frequency items. Finally, these results further support the need to investigate the individual differences among participants and combine corpus-based data with user-related data.

The result of each experiment described in this dissertation points to usage-based variation in mental representations of multimorphemic and multi-word units. The fact that the judgements revealed how prior experience with language determines the strength of memory representation is in line with findings from previous studies that have investigated the correlation between familiarity rating and processing and voice onset time (e.g. Verhagen (2020), Caldwell-Harris et al. (2012), Gardner et al. (1987)). By having the same participants perform a metalinguistic judgement task, as well as a range of behavioural tasks and analysing the data at the level of individual participants, it is possible to account for the individual differences which are stable across tasks, as suggested by the results of temporal, phonetic reduction, and strength of memory retrieval tasks. Familiarity scores in these experiments can be treated as frequency information retrieved from each participant's "mental corpus". Following this reasoning, the question emerges as to why frequency and familiarity significantly affect the processing, access, and production of linguistic items, especially the speech rate of the participants. According to Rutvik et al. (2018), there are several reasons why frequency and familiarity influence lexical retrieval during reading. Firstly, access to the meaning of familiar words requires less cognitive effort; it is more automatic because of repeated exposure. Second, high-frequency and thus familiar words have richer semantic representations, strengthening the exemplar itself according to exemplar theory. Third, these words have more associations, and according to the Hebbian learning theory, the simultaneous or near-simultaneous activation of multiple neuronal circuits promotes the formation of connections between them (Takashima & Bakker 2017). After accessing the word's lemma, the sound form of the words has to be encoded. At this stage, the connection between familiarity and neuromotor automaticity is visible. According to Bybee,

> the articulatory representation of words and sequences of words is made up of neuromotor routines. When sequences of neuromotor routines are repeated, their execution becomes more fluent (Bybee 2006: 715).

The results above have highlighted that generating the phonetic plan for words is connected to word familiarity. According to Levelt (1989), the phonetic plans for the syllables specify the articulatory gestures that later have to be executed. This plan can be characterised as a sequence of phones. However, one has to bear in mind that these phones are overlapping events. Each phone in the syllable is a temporal gesture that typically overlaps during its execution with other phones. Bybee (2002b) describes these overlaps as the result of the repetition of the neuromotor patterns. "The anticipatory overlap of gestures smoothes transitions between the phones, and unnecessary or extreme gestures decrease in magnitude or are omitted" (2002b: 268). One of the main claims of this dissertation is that the more quickly a participant was able to pronounce a word, the more synchronised the articulatory organs during the pronunciation of that particular word are. The synchrony assumes that the synopsis between the responsible neurons that transmit the command to the speech organs is quicker. This process can be interpreted as the result of entrenchment, which is reflected in the automatic production of the selected items. If the word is pronounced more quickly, this presumes that it has a strong memory representation. In contrast, if it is pronounced slowly or more slowly than another word, it could mean that the synaptic connection between neurons is established for the first time, and the articulatory organs are not yet synchronised. The lack of synchrony was manifested by the fact that the participants were reading the unknown words syllable-wise, using the phoneme-to-grapheme route during reading instead of the direct route.

In conclusion, the results of the analysis of familiarity ratings as a direct measure of speakers' experience with the language and the strength of memory representations demonstrated the importance of paying closer attention to the usage-based variation in behavioural experiments. Eliciting familiarity ratings from participants who also participate in different behavioural tasks provides an additional perspective. It can help us, as researchers, to better understand how language is stored and processed. Selecting participants with different social backgrounds is not a fine-grained enough predictor of previous language experience since not only do groups as a whole vary in their language history, but each member of the group might have a different experience with the language. The analysis of the behaviour of each individual provides unique explanatory power that can indicate which factors are significant predictors of the group's performance and which predictors allow us to analyse the variation observed among participants. In other words, the observed variation is not random noise but a valuable source of information.

9.2 Finding 2

A tendency that emerged from the literature review is that there is competition between different linguists to settle the issue of which linguistic predictor has the highest explanatory power: frequency, transitional probability, or any other corpus-based measures. With the arrival of large corpora, the study of frequency effect has boomed. Moreover, these corpora were used to develop new measures of word attributes. A range of statistical methods was developed to determine which of the many lexical variables would best predict the results of the lexical decision and word-naming tasks (Caldwell-Harris 2021). The first part of this revolution was moving away from pure frequency-based measures and toward the probabilistic view of language. The development of statistical and machine-learning technologies allowed researchers to calculate increasingly complex measures (like entropy or informativity) and fit them into complex statistical models to show which of them has the highest explanatory power. These linguistic factors were often compared pairwise, e.g., frequency and transitional probability or frequency and contextual diversity. The current study aimed to compare them all and investigate their possible interaction.

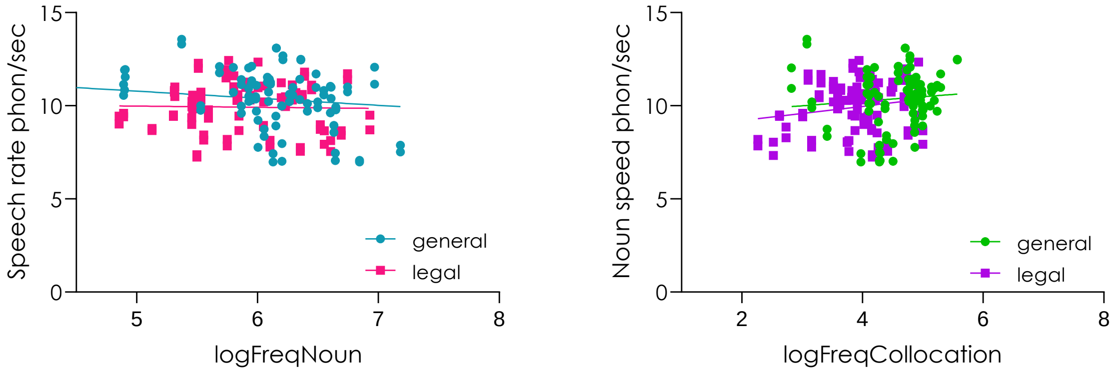
The four experiments aimed to highlight that it is challenging to separate the effect of the different corpus-based measures as they rely on frequency in one form or another. Thus, analysing their combined effect and investigating their interaction is more fruitful. One of the central claims highlighted in the theoretical framework of this dissertation is that entrenchment stands on three pillars, and input alone is not sufficient to achieve entrenchment since language users rely on a range of cognitive skills that allow them to register the input, encode it into memory and automatise the language processing and production of these stored elements. In order to validate this claim, most of the linguistic and cognitive predictors that are associated with entrenchment were incorporated into the data analysis. Moreover, the experiments aimed to underline the importance of combining linguistic factors with social and cognitive factors to provide empirical validation for the processes associated with entrenchment.

The input to which language users are exposed is rich in metalinguistic data; some words occur more frequently than others, some adjectives are more likely to be followed by a particular noun, and some seem to appear in various contexts. Some language users might focus more on the frequency of the words, others more on transitional probabilities, and still others might attend to the different contexts in which the words occur. For this reason, it is doubtful that one of these factors could explain alone the observed variation. Furthermore, it is suggested that even if language users pay more attention to some of these factors, it is doubtful that they completely neglect the others. In other words, language users might realise that the word *foreseeable* is highly likely to be followed by the word *future* (approximately 83% of the time, according to the English Web 2020 corpus). Moreover, they might also register that it is a frequent collocation and in which context it is appropriate to use and likely to be observed. Thus, due to the rich metalinguistic data that we register about the stored linguistic items, we are likely to rely on them to a higher or lesser extent depending on the communicative situation or whether we are learning or retrieving the items.

9.2.1 The effect of the linguistic factors, such as frequency and transitional probability, is task, paradigm and person-dependent.

A detailed analysis of the different linguistic factors' effect on temporal/phonetic reduction and strength of memory representation revealed that none of the factors demonstrated a consistent effect across the different experiments and tasks, except familiarity, which has already been discussed in the previous findings.

First, the effect of these factors on temporal reduction will be discussed in detail, followed by phonetic reduction and accuracy of memory recall. As Figure 39 below shows, none of the corpus-based measures has an exceptionally significant correlation with the rate of temporal reduction.



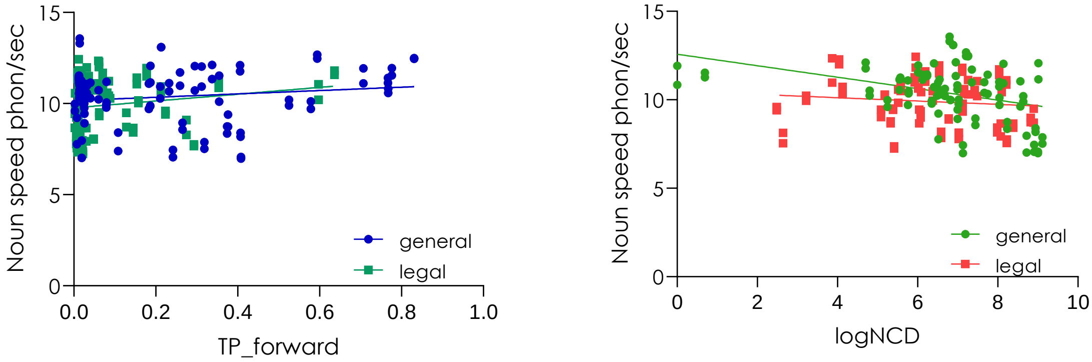


Figure 39. The effect of noun frequency, collocation frequency, forward transitional probability, and noun contextual diversity on the rate of temporal reduction in experiment IV

Figure 39 demonstrates the frequency of the collocation, and the forward transitional probability seem to have the most significant effect on the rate of temporal reduction. However, based on the existing research, their effect is less powerful than expected. Surprisingly, NCD and the frequency of the noun show a reverse effect; thus, the higher the frequency or the CD of the noun, the more slowly they were pronounced. A probable reason for the reverse effect NCD has on the rate of temporal reduction is that the higher the noun's CD, the more information is subconsciously activated about the noun. In other words, the mental exemplar of the noun is rich in diverse information, and more information is retrieved upon encountering the word, which slows down the access and pronunciation of the word.

A further inspection of the other two experiments has revealed that the effects of the linguistic predictors are not stable across the different paradigms. Figure 40 below demonstrates that when the same linguistic predictors' effect is analysed on the phonetic reduction rate, then the collocations' frequency alone shows a different effect on the legal and general collocations. It has a positive effect on reducing the prenasal schwa in legal words and a negative one on reducing the general words.

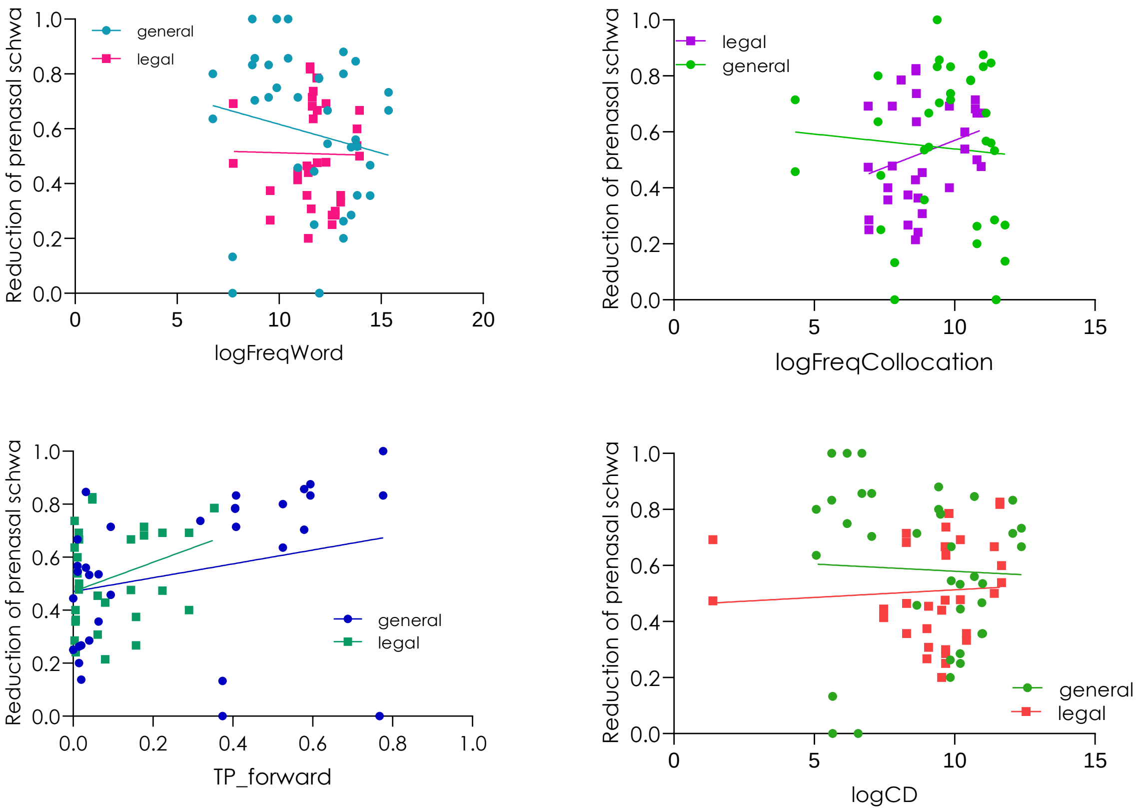


Figure 40. The effect of noun frequency, collocation frequency, forward transitional probability, and noun contextual diversity on the rate of phonetic reduction in Experiment IV

The forward transitional probability is the only variable that shows the same tendency across the two experiments. The CD[98] of the selected words does not significantly affect the rate of phonetic reduction, and the frequency of the noun demonstrates the adverse effect, but only for the general words. This last finding is not as surprising as it might sound. Song (2013) found that low-frequency words were more likely to include a deleted schwa than high-frequency words. She claims that there is no significant relationship between the frequency of the word and the schwa elision rate. Raymond et al. (2006) found mixed effects while analysing the word-internal /t, d/ elision. The frequency of the word containing a word-internal alveolar stop did not predict deletion in the overall dataset, nor did it predict deletion in function or content words when these were tested separately. However, when the four subsets of content words defined by syllable position and preceding segment class were tested separately, the frequency of the token word did predict deletion but only for onset tokens in V_ contexts of content words (Raymond et al. 2006).

[98] The list of words that were selected for the analysis included not only nouns, but also adjectives and verbs. For this reason, these words' contextual diversity was extracted from *Die Zeit* corpus.

The analysis of the strength of memory representations obtained from the sentence recall task further supports the claim that the effect of the linguistic variables is not stable across the different paradigms, and the participants are likely to rely on distinct factors depending on the task and communicative situation. Figure 41 demonstrates that the frequency of the noun and the NCD had a facilitatory effect on recall accuracy. However, in the previous two experiments, they showed a reverse effect.

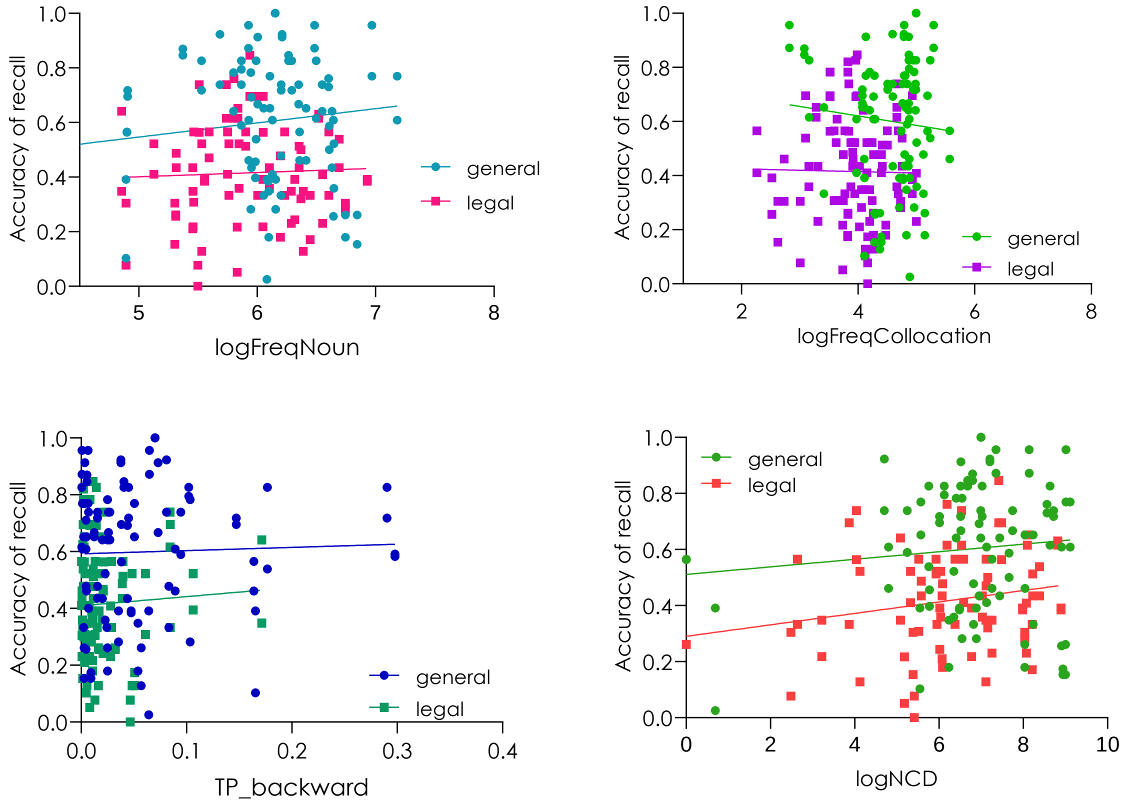


Figure 41. The effect of noun frequency, collocation frequency, forward transitional probability, and noun contextual diversity on the accuracy of recall in Experiment IV

Furthermore, the forward transitional probability was substituted by backward transitional probability because, in the statistical model analysing the recall accuracy, only backward transitional probability has reached the significance threshold. Backward transitional probability did not reach the significance threshold in any of the other experiments or paradigms.

Moreover, the analysis provided here is based upon experiment IV, as this was the only experiment that included all four measures across the three different paradigms (temporal reduction, phonetic reduction, and strength of memory representation). Table 21 below summarises the effect of the linguistic variables across the different experiments and paradigms. The "x" in the different cells indicates that the particular predictor was not part of the analysis.

Table 21 is revealing from many perspectives. First of all, it shows that the tendency observed and highlighted in experiment IV is not unique. It is observed across the different experiments and paradigms. There is no single linguistic predictor that is significant across all the experiments. This finding, however, does not mean that these linguistic factors do not facilitate the process of entrenchment. The results imply that distinct factors facilitate the process depending on the communicative situation. For the analysis of temporal reduction, frequency seems to have a facilitatory effect. However, this effect is not present in phonetic reduction. On the other hand, backward transitional probability is only a significant predictor of memory recall. Second, the result shows that the effect of these linguistic variables has to be tested in different paradigms, not only in lexical decision tasks or naming latencies, to understand their role in the different language-related processes.

The review of the literature on this topic shows a remarkably similar picture. Numerous studies have compared the effects of various linguistic predictors on performance, but none of them reached a consensus. Most of the seminal studies focused on comparing frequency and a further measure, e.g., contextual diversity or transitional probability using lexical decision tasks or naming latencies. However, very little research has been done on determining which measures work best under which circumstances and even fewer, if any, that incorporated a range of these measures into their analysis and compared their effect across different tasks.

Brysbaert & New (2009) found that CD was a better predictor of single-word processing times than word frequency was. Baayen (2010) had similar findings using subtitle databases. Jones et al. (2012) went even further and used semantic distinctiveness, word frequency, and contextual distinctiveness to predict humans' recognition accuracy of spoken language words with varying noise levels. Semantic distinctiveness predicted more variance in response times than the other two variables. The three studies provided empirical evidence for CD's significant effect on word recognition. Surprisingly, none of the experiments conducted within this dissertation could confirm this effect. It could be because the experiments used different methodologies and focused not only on single words but also word combinations.

Table 21. Summary of the effect of the linguistic predictors across experiments and paradigms. The significant effects are in bold

Experiment		**FreqWord**	**FreqColloc**	**Forward TP**	**Backward TP**	**CD**
Temporal reduction	Exp I	**-0.023451[99] (p = 0.03)**	x	x	x	x
	Exp II	1.1286 (p = 0.288)	**-3.223e-02[100](p =0.002)**	**-5.689e-02 (p = 0.001)**	x	x
	Exp III	**-4.554e-02 (p < 0.001)**	4.917e-04 (p = 0.752)	-1.595e-04 (p = 0.57)	x	x
	Exp IV[101]	**0.059185 (p = 0.107)**	**2.153e-01 (p = 0.13)**	**4.244e-01 (p = 0.01)**	0.015314 (p = 0.69)	**-0.4479 (p <0.001)[102]**
	Exp II (schwa)	0.07763 (p = 0.84)	0.09746 (p = 0.59)	0.37142 (p = 0.51)	x	x
Phonetic reduction	Exp II (/t/ elision)	**-5.70746 (p < 0.001)**	x	x	x	x
	Exp IV (schwa)	-0.188189 (p = 0.59)	0.159365 (p = 0.53)	**0.66987 (p < 0.001)**	0.01696 (p = 0.92)	-0.03027 (p =0.83)
Recall	Exp III	-0.083465 (p = 0.52)	0.013755 (p = 0.83)	-0.147830 (p = 0.29)	x	x
	Exp IV	**0.49098 (p = 0.002)**	-0.35894 (p = 0.21)	0.19585 (p = 0.42)	**0.71808 (p < 0.001)**	0.1030 (p = 0.54)

[99] In experiments I, II, and III the raw pronunciation length was used as the dependent variables. Thus, the - sign indicates that the higher the frequency was, the faster the participant pronounced the target word.

[100] In experiment II, the stimuli were multimorphemic compounds, FreqWord refers to the frequency of the base, while FreqColloc refers to the frequency of the compound, Forward TP consequently, is the likelihood of the root followed by the ending.

[101] In experiment IV, the dependent variable was measured phon/sec, the linguistic predictors were expected to show a positive correlation. FreqColloc and Forward TP lost their significance once the interaction between them was added to the model.

[102] Contextual diversity is significant, but the direction of the effect is opposite of what was expected.

The lack of consensus suggests that there is a need to investigate how the effect of CD varies across different tasks and whether it is significant only for single words or whether its effect can be extended to more complex units, like collocations or even constructions. A further question that has to be addressed is whether these predictors demonstrate a different effect during the learning stage than during the recall and access stage. A recent study by Pagán & Nation (2019) examined whether variations in contextual diversity, spacing, and retrieval practice influenced how well adults learned new words from reading experience. Eye movements were recorded as adults read novel words embedded in sentences. Adults learned words faster when frequency (here, it refers to the number of repetitions in the text) was high, even if words were clumped together in the same context. In other words, frequency facilitated word learning to a greater extent than contextual diversity. Another study by Mak et al. (2021) obtained similar results: words experienced in the low-diversity condition showed a better learning rate than words in high-diversity conditions. They attributed these findings to "anchoring", a process of stabilising novel word representations by securing them onto a familiar topic in long-term memory.

All these findings imply that contextual diversity has the most significant effect on single-word recognition. This assumption aligns with the claim that the effect of these linguistic predictors is task- and situation-dependent, and they can compensate for each other. Moreover, as Caldwell-Harris (2021) suggested, a possible explanation for these different results is that the facilitatory effect of contextual diversity and word frequency is due to the different underlying mechanisms. This suggestion is based on empirical evidence that comes from the study of Vergara-Martínez et al. (2017). Their study observed different ERP patterns for high-CD versus high-frequency words. High-CD words had a larger ERP negative component - ranging from 225 to 325 ms post-stimulus onset - than low-CD words did. This result goes in the opposite direction of the standard N400 word frequency effects, which have low negative amplitudes. The explanation they have provided for the opposite effect of frequency and contextual diversity is that a word's memory trace is strengthened by merely repeating words, and access to it is made easier over time. With each occurrence, the memory trace becomes stronger, enhancing the efficiency of access to it on subsequent occasions. By contrast, contextual diversity reveals the ways in which words are represented, specifically, the variation in meaning that is enhanced by the numerous contexts within which a word is observed.

To my knowledge, no studies have compared the underlying processes that determine the facilitatory effect of transitional probability. However, it can be assumed that just as frequency and CD have different underlying processes, transitional probability also relies on a separate process, contributing to the mental representations in a unique way. Based upon the theoretical framework outlined in this dissertation, it can be assumed that transitional probability facilitates the emergence of different patterns and helps to link the word to other words and, in this way, unite them into one network. Due to the linear unfolding of the language both in production and processing, these connected words show a facilitatory effect both on the naming latencies and processing times.

A similar tendency is observed in the studies that have compared the effect of transitional probability with word frequency. A study by Moers et al. (2017) investigated the effect of frequency and transitional probability in reading aloud speech and how it varied in younger and older speakers. Their participants fell into three groups: children (8-12), young adolescents (12-18) and older participants (62-95). Their methodology was very similar to that used in experiments III and IV. The results revealed that transitional probability influenced the processing of words in context, beyond the well-established effect of word frequency, across the entire age range. However, the results also indicated that age groups differ in the size of TP effects, with older adults having smaller TP effects than adolescent readers. The contrary was observed in the study by McConnell & Blumenthal-Dramé (2021). The results of their self-paced reading task showed that age enhanced the speeding-up effects associated with transition probability. They attributed these results to the fact that older adults rely more heavily on predictive processing and have more entrenched probabilistic knowledge. None of the experiments within this dissertation that used transitional probability has observed such interaction with age in any of the paradigms (temporal/phonetic reduction and strength of memory recall). However, it has to be emphasised that none of the experiments covered such a wide age range as those experiments mentioned above. Furthermore, the two studies by McConnell & Blumenthal-Dramé (2021, 2019) revealed that backward transitional probability was a more significant predictor of the processing times than frequency or forward transitional probability. In the present study, forward transitional probability significantly influenced the rate of temporal and phonetic reduction and backward transitional probability the accuracy of memory retrieval. Endress & Langus (2017) argue that role of transitional probabilities might not

be to memorise items per se but rather to prepare learners for memorising recurring items once they are presented in subsequent learning situations with richer cues.
To further highlight the lack of any consensus between studies that compare different linguistic predictors in different tasks, the study by Lorenz & Tizón-Couto (2019) has to be mentioned. Their study investigated a less researched area: how transitional probability influences language comprehension of full and reduced forms. The results obtained suggest that transitional probability is a facilitating effect on reduced but not full forms. On the other hand, the recognition of the full forms was facilitated by frequency. Their follow-up[103] study used the same methodology and stimuli and tested L2 learners in their recognition times of full and reduced forms. Surprisingly, neither the recognition of reduced nor of full forms profited from higher transitional probabilities.
In conclusion, the only common thing about these studies is that they emphasise that the different linguistic predictors used to model the participants' performance are task, communicative situation and age-dependent. When multi-word units are investigated, factors that predict perfectly single-word recognition and naming latencies lose their effect. It is due to the fact that the context in which a word may occur plays a role, making single-word factors such as frequency less important. Furthermore, what is a significant predictor for processing times does not always facilitate learning and vice versa. A further important implication that is borne out of the experiments conducted within this dissertation and the discussion of previous research above is that individual differences cannot be neglected. At certain stages of language development, language users may rely more on frequency than transitional probability or contextual diversity, while at another stage, they may rely on completely different factors. Moreover, these predictors can compensate for each other. If word frequency is low, but the transitional probability is high between the processed words, these words may not show any production or processing disadvantages, as one component would compensate for the other.

[103] At the moment of writing this dissertation no official publication was available. The results of the study were presented during the DGKL 9 conference and one of the authors, David Tizón-Couto, shared their presentation materials with me.

9.2.2 The interaction between the different linguistic, social and cognitive factors has a cumulative and compensatory effect on the performance of the participants.

Very few studies have examined the interaction between linguistic, social, and cognitive factors. The interaction between the linguistic and other predictors can have two possible outcomes. First, each factor may play an equal role in the rate of temporal, phonetic recall, and accuracy of recall. Another possibility is that some factors are more influential than others, and a more decisive factor might nullify the influence of a weaker factor. As a word or collocation can be highly probable according to one factor (e.g., lexical frequency, contextual diversity) and highly improbable according to another factor (e.g., conditional probability based upon the preceding word), these factors can work in opposite directions. Thus, they can potentially cancel each other out. However, if they are both working in the same direction, i.e., if they show a facilitatory effect, their effects may be additive, multiplicative, or one effect may be much larger than the other, hiding the effect of the weaker factor. Furthermore, it is also possible that language users pay more attention to some of these predictors and rely upon them more when engaged in language comprehension and production.

9.2.2.1 Linguistic factors

The analysis of the different experiments has revealed several significant interactions between the linguistic variables. One of the surprising interactions that emerged is that between transitional probability and item type in the analysis of temporal reduction. This interaction was observed both in experiments III and IV. The interaction observed is visualised in Figure 42 below.

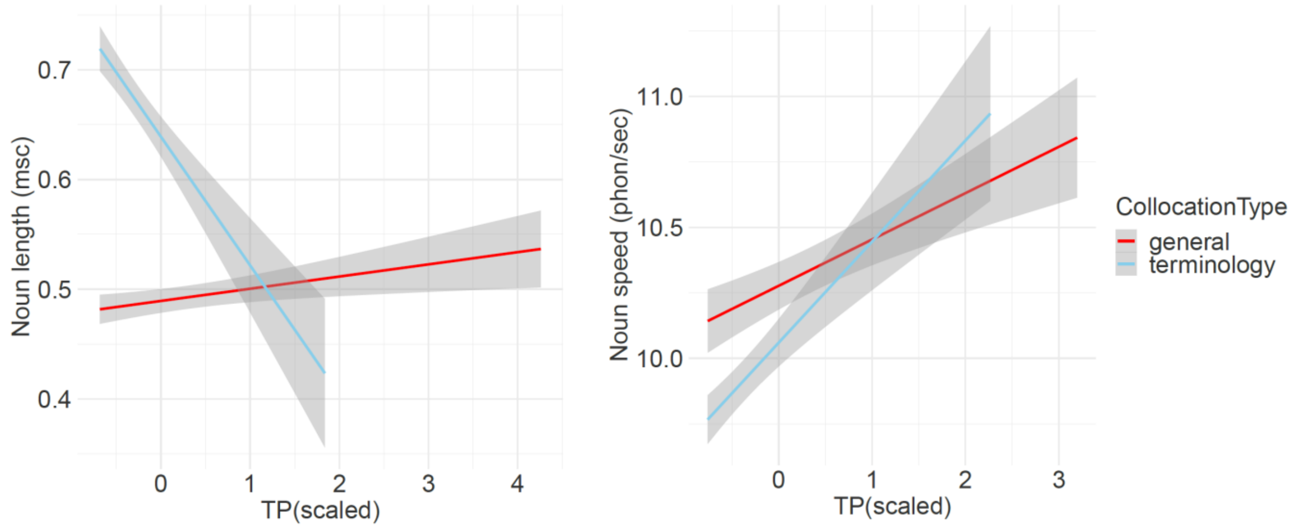
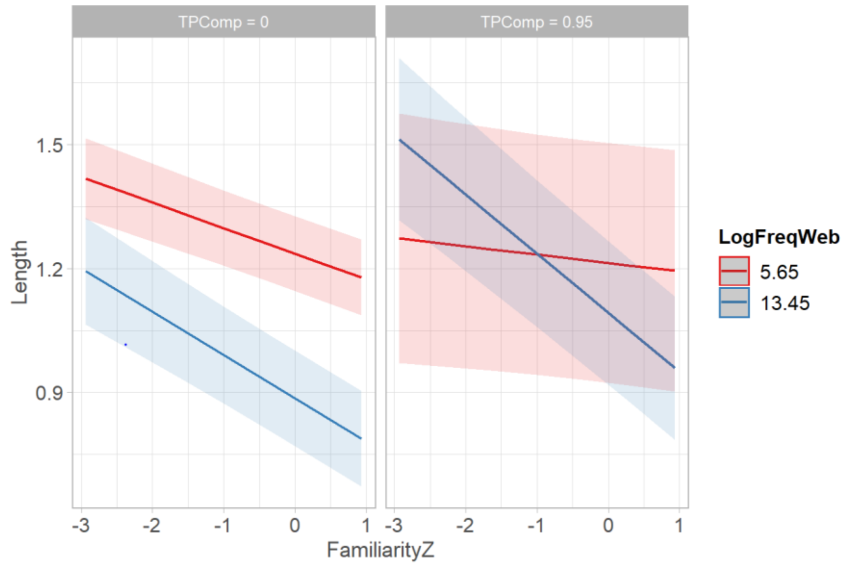


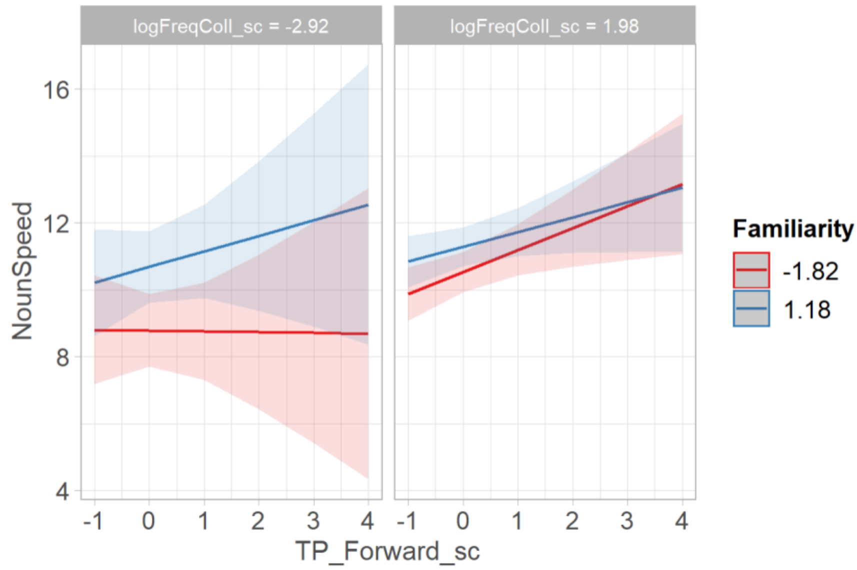
Figure 42. Interaction between item type and transitional probability. The figure on the left is from experiment III (it shows the noun pronunciation length in milliseconds), and the one on the right is from experiment IV (it shows noun speed in phon/sec)

Figure 42 reveals that higher transitional probabilities more significantly affected legal collocations. A possible explanation for the results observed is that legal collocations were less frequent and familiar to the participants, and they benefited more from contextual predictability when they needed it the most. Higher transitional probabilities compensated for the lack of familiarity and higher frequency. If listeners have stored the collocations as chunks and built-up expectations about the upcoming words, they need less time to activate and produce the word. These results can be compared to those obtained by Lorenz & Tizón-Couto (2019), where reduced forms benefited more from higher transitional probabilities than the full forms. The fact that transitional probabilities correlate with item type also implies that language users rely more on their built-up expectations when the input has a higher cognitive load. In experiment IV, contrary to Lorenz & Tizón-Couto's (2019) results, transitional probabilities significantly affected both legal and general collocations, but legal collocations benefitted more. However, in experiment III, transitional probability showed a reversed effect on the general collocations, i.e., higher TP slowed down the pronunciation rate. It is possible that this reverse effect emerged due to the small sample size.

A further interaction that has been observed in more than one experiment is the three-way interaction between familiarity, frequency, and transitional probability.



a)



b)

Figure 43. The three-way interaction between familiarity, frequency, and transitional probability. Figure a) represents experiment II, and figure b) represents experiment IV. Figure a) shows the y axis in milliseconds and in b) in phon/sec

Figure 43 demonstrates a logical and expected interaction: the higher the frequency, familiarity, and transitional probability are, the faster participants can pronounce the word. Thus, their effect did not cancel each other out but was cumulative. Nevertheless, to my knowledge, no studies have investigated this interaction. Most of the studies that included more than one linguistic predictor in their analysis aimed to determine which of them was more significant and did not consider their mutual effect on the participants' performance. Moreover, figure 43 supports the claim made above that these linguistic factors can compensate for each other. Figure 43 a) shows that when the target item's transitional probability is low, there is no significant difference between low- and high-frequency words. However, when the transitional probability is high and the collocation itself is frequent, their effect accumulates, resulting in an even faster speech rate. Figure 43 b), on the other hand, demonstrates that when both frequency and transitional probability are low, the participant's familiarity with the target item can compensate for it and lead to faster pronunciation. Interaction between the linguistic predictors was also observed in modelling the strength of memory recall. The three-way interaction among word length, familiarity and noun speed significantly impacted recall accuracy. Noun speed can be regarded as a cognitive factor, as there are individual differences among the participants regarding their speech rate.

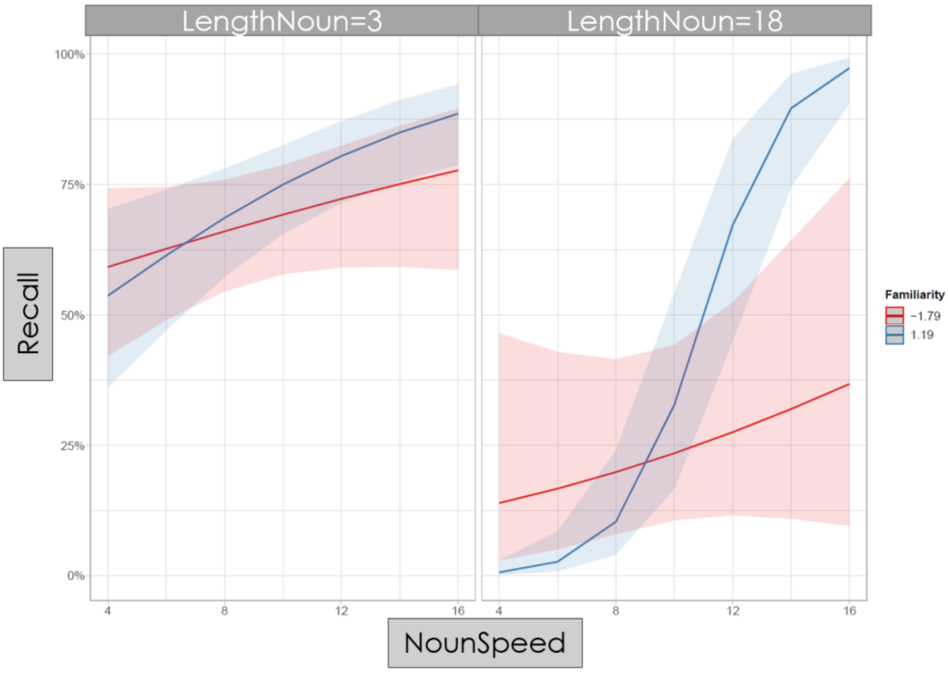


Figure 44. The three-way interaction among word length, familiarity, and noun speed in the modelling of the strength of memory retrieval

Figure 44 reveals that longer words in the recall task benefited significantly from the higher familiarity rating and higher noun speed. However, higher familiarity ratings also facilitated the pronunciation of shorter words, not as significantly as long words. This can further imply that familiarity compensated for word frequency since Zipf's law states that longer words tend to be less frequent. The study on variability in word duration by Baker & Bradlow (2009) found a significant three-way interaction among lexical frequency, discourse mention, and speaking style. High-frequency words exhibited more second-mention reduction than low-frequency words in plain but not in clear lab speech. McConnell & Blumenthal-Dramé (2021) study analysed the four-way interaction between age, reading experience, position, and forward/backward transitional probability in their self-paced reading study. However, no three- or four-way interactions were statistically significant. In the current experiment, interactions were observed among the linguistic factors and among the social factors, but no interaction was detected across the different factors, i.e., between linguistic and cognitive or social and cognitive.

9.2.2.2 Social factors

Besides the interaction observed among the linguistic predictors, a significant interaction between the social variables in experiment II was also observed, namely, the interaction between age, gender[104], and occupation. The interaction is illustrated in Figure 45.

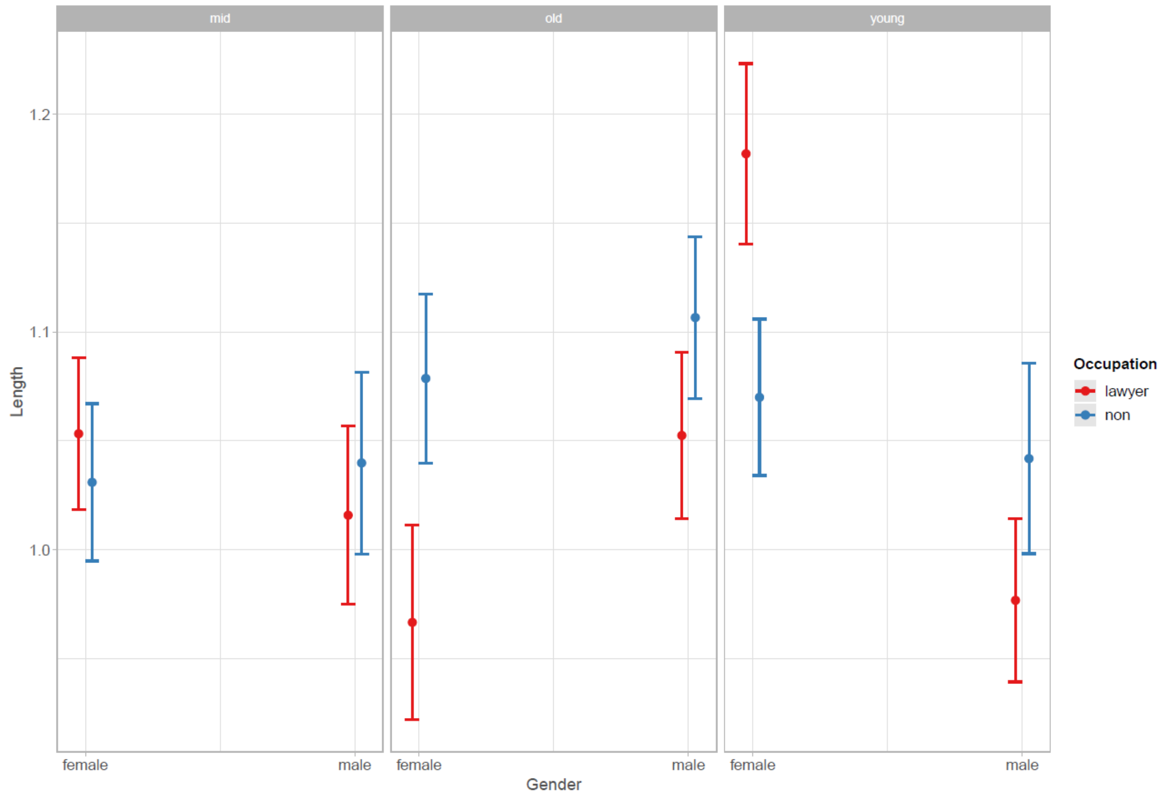


Figure 45. The interaction between the social variables in Experiment II

This was the only experiment where the age difference was significant enough to add it to the interaction. Figure 45 shows that older female law participants could pronounce the target word the fastest. This contradicts the expectations, as the older male law participants had the most experience with legal words. Moreover, the young

[104] Gender is treated as a binary categorical variable because none of the participants indicated otherwise in the personal background questionnaire. The study does not aim to make any claims about the language or cognitive capacities of males and females. It only demonstrates the interaction observed.

female law participants showed the slowest pronunciation. The interaction observed here is likely due to an unbalanced distribution of participants across age groups and gender. The effect of the social predictors displayed across the different tasks and paradigms is summarised in Table 22 below.

Table 22. Summary[105] of the effect of the social factors across the experiments and paradigms. Significant predictors are in bold type

Experiment		Age	Gender:M	Occupation: Non
Temporal reduction	Exp I	0.00072 (p = 0.95)	0.09276 (p = 0.55)	**-0.840 (p = 0.002)**
	Exp II	-6.719e-04 (p = 0.31)	**-3.74e-02 (p = 0.004)**	**-2.23e-02 (p = 0.012)**
	Exp III	4.090e-04 (p = 0.71)	-1.265e-02 (p = 0.11)	**3.660e-02 (p < 0.001)**
	Exp IV[106]	-1.011e-02 (p = 0.25)	**9.731e-0.2 (p = 0.05)**	-8.127e-02 (p = 0.10)
Phonetic reduction	Exp II (schwa)	**0.033 (p=0.01)**	-0.01 (p=0.94)	**-0.5491 (p = 0.053)**
	Exp II (/t/ elision)	**0.05376 (p = 0.01)**	-0.1182 (0.74)	0.5027 (p = 0.19)
	Exp IV (schwa)	-0.0522 (p = 0.23)	0.0217 (p = 0.93)	0.1119 (p = 0.64)
Recall	Exp III	**0.08615 (p = 0.03)**	0.1180 (p = 0.705)	0.0576 (p = 0.89)
	Exp IV	0.026 (p = 0.01)[107]	**0.1852 (p = 0.002)**	**- 0.3697 (p < 0.001)**

Table 22 demonstrates that the effect of the social factors is not consistent. Only occupation seems to be significant in most cases and across the different paradigms, as was expected. It was not one of the main objectives of the experiment to outline the differences between the gender, and for this reason, during the recruitment, equal distribution between the genders was not achieved. The results obtained are possibly due to the fact that the different genders were not equally represented, and older participants might have had more experience with legal words than the younger participants. The previous literature also shows a mixed effect concerning the role gender might play in phonetic reduction. Among the researchers who investigated the connection between age, gender, and phonetic reduction are Guy (1992), Neu

[105] Due to the fact that not all the final models included all the social factors, data was taken from the global models. This could cause certain deviations from the data reported in the tables of the final models in the previous section.

[106] In experiment IV, the dependent variable was measured phon/sec, the linguistic predictors were expected to show a positive correlation. FreqColloc and Forward TP lost their significance once the interaction between them was added to the model

[107] Age has lost its significance in the final model

(1980), Zue & Laferriere (1979), Raymond et al. (2006), Brückl & Sendlmeier (2003), Skoog et al. (2015). Both Guy (1992) and Raymond et al. (2006) found a negative correlation between /t, d/ deletion and age; thus, younger participants are more likely to delete /t,d/ sounds in speech. Neu (1980) and Zue & Laferriere (1979) investigated the gender differences concerning the deletion processes. Neu's study (1980) revealed differences among the male and female participants concerning the sensitivity towards the phonological environment: men showed a higher tendency to delete the word-final alveolar stop when it was preceded by a sibilant rather than by nasals or stops.

On the other hand, females did not demonstrate this sensitivity toward the phonological environment in which the sound occurred. Interestingly, the study by Zue & Laferriere (1979) came to a different conclusion. Their results indicated that females produced more reduced forms in a nasal-flapping environment than men. Finally, the results obtained by Raymond et al. (2006) revealed no effect of gender on the deletion processes. Byrd's (1994) study revealed that males actually speak somewhat faster than females. Yuan, Jiahong, et al. (2006) have obtained small but significant differences between the speech rate of males and females, confirming the results of Byrd (1994) that male participants speak slightly faster.

The review of the previous literature and the analysis of the four experiments show a similar tendency, which implies that the type of the task can greatly influence which factors reach a significance level. Moreover, it is possible that some hidden factors, such as experience with the language or the participants' cognitive capacities, can indeed influence their performance to a larger extent than their age or gender.

9.2.2.3 Cognitive factors

This assumption is confirmed by the subsequent finding. All the experiments, except one, aimed to incorporate some measures of the participants' cognitive skills. The theoretical framework outlined at the beginning of this dissertation emphasised that language users' different cognitive skills help them extract information from the input and entrench it. With the development of the test batteries, more and more complex measures of the participants' cognitive skills were added to the experiments. In experiment II, only the participants' reading skills were measured. Reading skill, strictly speaking, is not a cognitive skill. It was used to obtain a vague impression of the participants' reading habits. It was assumed that the faster participants read, the more likely it is that they spend more time reading, which increases the chance that they have already witnessed the target words and might have entrenched them. Reading habits were measured using the self-paced reading task in all the experiments. In the final experiment, reading habits were re-classified as verbal processing skills. The rationale behind this was that the previous experiments (experiments II and III) had shown a significant correlation between the rate of temporal reduction and phonetic reduction and the time participants spent reading the selected texts. Their speed was able to reveal how fast they processed written information. All the experiments required participants to read aloud words and sentences. For this reason, it is not surprising that those participants who were faster in silent reading were also able to display more fluent and faster pronunciation whilst reading aloud. The processes that underlie silent reading and reading aloud are the same, except that reading aloud also requires the activation of the motor cortex to pronounce the words. The models focusing on temporal reduction and strength of memory representation revealed that reading skills, later referred to as verbal processing speed, rendered a significant effect.

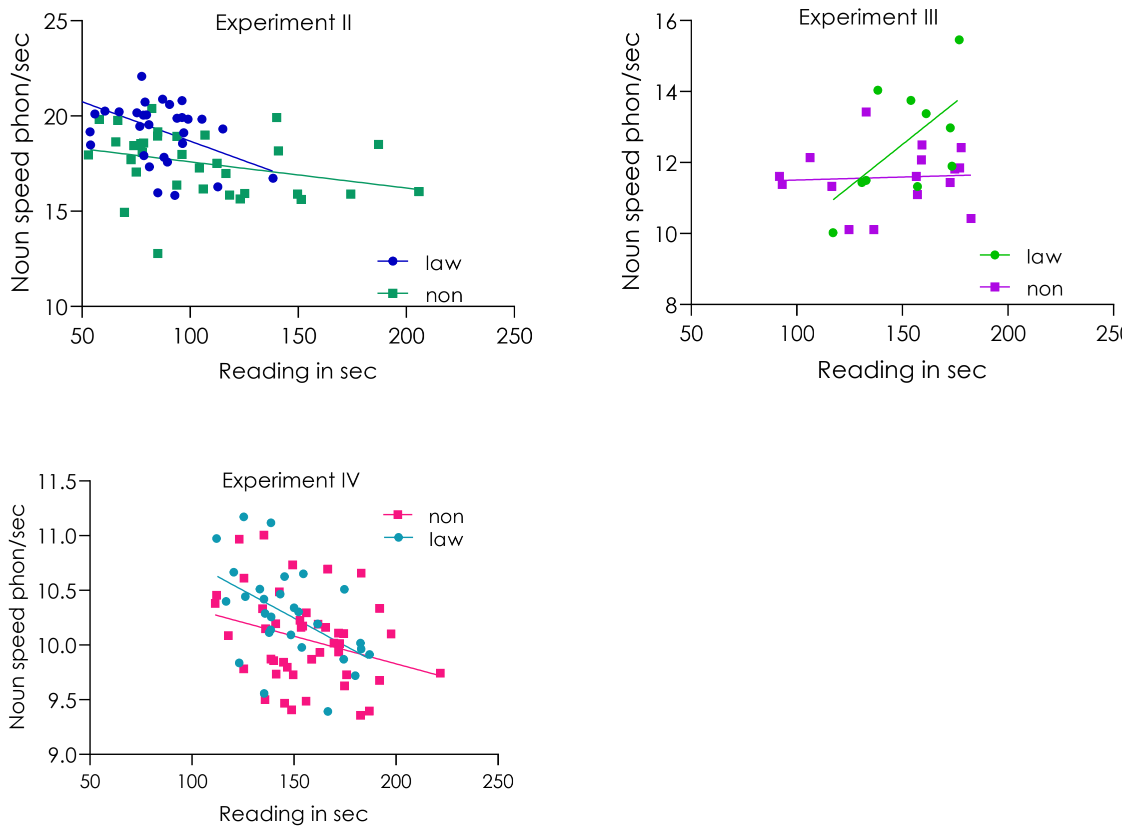


Figure 46. The effect of reading speed on noun speed across the experiments

Figure 46 shows that, in 2 out of the 3 experiments, reading significantly affected the selected items' temporal reduction. The reason why, in experiment III, the tendency is not observed is due to the very small data set. It was highlighted that, unfortunately, temporal reduction in experiment III could not be used to draw any conclusions about the effect of the different linguistic, social, and cognitive factors on the temporal reduction rate. To my knowledge, no studies have investigated the participants' reading habits or verbal-processing skills on the rate of temporal reduction. For this reason, the current results can neither be supported nor questioned based on the existing literature. Further research is required to understand whether this correlation is observed only in this data set or whether verbal processing speed can indicate something about processes that underlie temporal reduction.

Reading had a reverse effect in the models analysing the strength of memory representation. Thus, longer reading times lead to a more accurate recall of the target items. This reversed effect indicates that those participants who processed the texts more slowly paid more attention to them, and this meticulous attention led to a better memorisation of the target collocations.

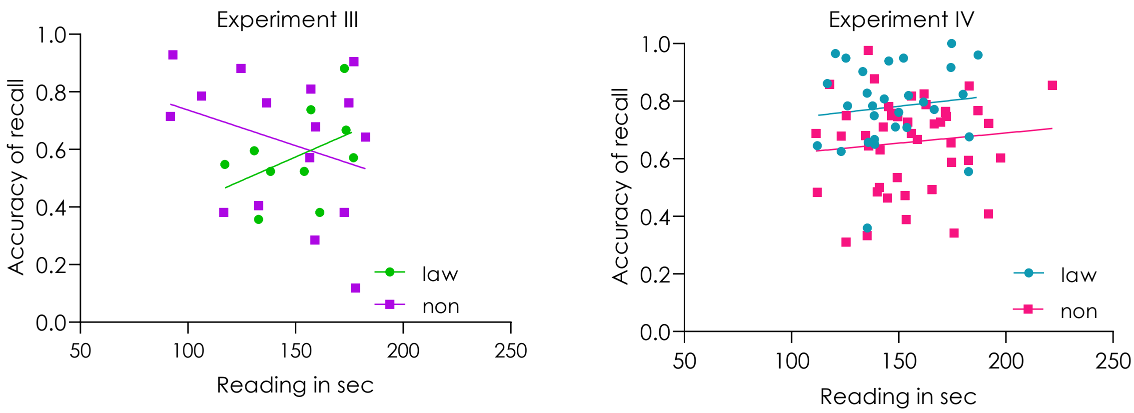


Figure 47. The effect of reading on the strength of memory representation

To a certain degree, this attentive reading manner can be regarded as a cognitive skill. More attentive language users are more likely to acquire new words, collocations, and grammar patterns faster.

In experiment III, reading had a different effect on the law and non-law groups. A possible explanation for this effect is, as has already been mentioned in the context of temporal reduction, the small sample size or the fact that there were two different texts, a legal and general text, and that they were taken together when plotting the graph. Legal texts were too difficult for the non-law participants, which could also influence the results obtained.

Besides reading, the participants' memory capacity and learning and non-verbal processing skills were measured, and their effect on recall accuracy was investigated. Memory capacity was the focus of experiment III and the learning skills of experiment IV. Non-verbal processing skill was part of both experiments. It seems logical that the higher the participants' memory capacity, the more accurately they would recall the target collocations. Memory capacity was measured using three different test batteries: number span, word span and reading accuracy.

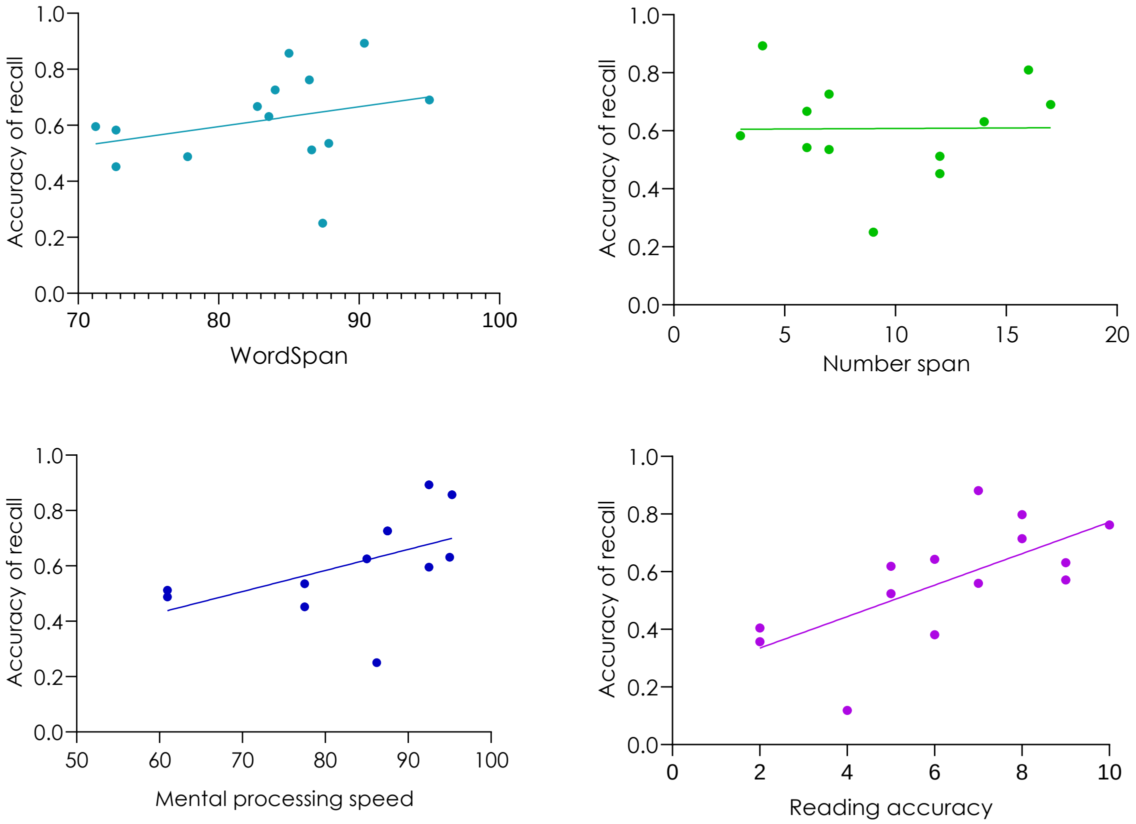


Figure 48. The effect of the cognitive measures on the accuracy of recall in Experiment III

Of the four cognitive measures, two significantly affected the accuracy of memory recall. Contrary to expectations, neither word nor number span significantly correlated with recall accuracy. Reading accuracy, as expected, showed a positive correlation. This expectation stems from the analysis of the effect of reading on recall accuracy. Participants who read more carefully were able to remember more collocations. Surprisingly, mental processing speed showed a significant correlation with recall accuracy. Only very few studies have investigated the effect of processing speed on verbal memory. These studies stem from clinical investigations that focus on schizophrenia (McDowd et al. 2011), hearing impairment (Kestens et al. 2021) and balancing skills (Blodgett et al. 2021). These studies established a positive correlation between verbal memory and processing speed. However, these studies cannot be directly compared to those presented in this research as all the participants were without any kind of disabilities and were younger than those discussed in the three studies. Thus, further research with healthy participants is needed to verify the correlation observed between the two phenomena. Due to the insignificant effect of the memory capacities tests in experiment III, they were replaced by measures of learning skills in experiment IV. Figure 49 below represents the effect of the cognitive factors on the accuracy of recall in experiment IV.

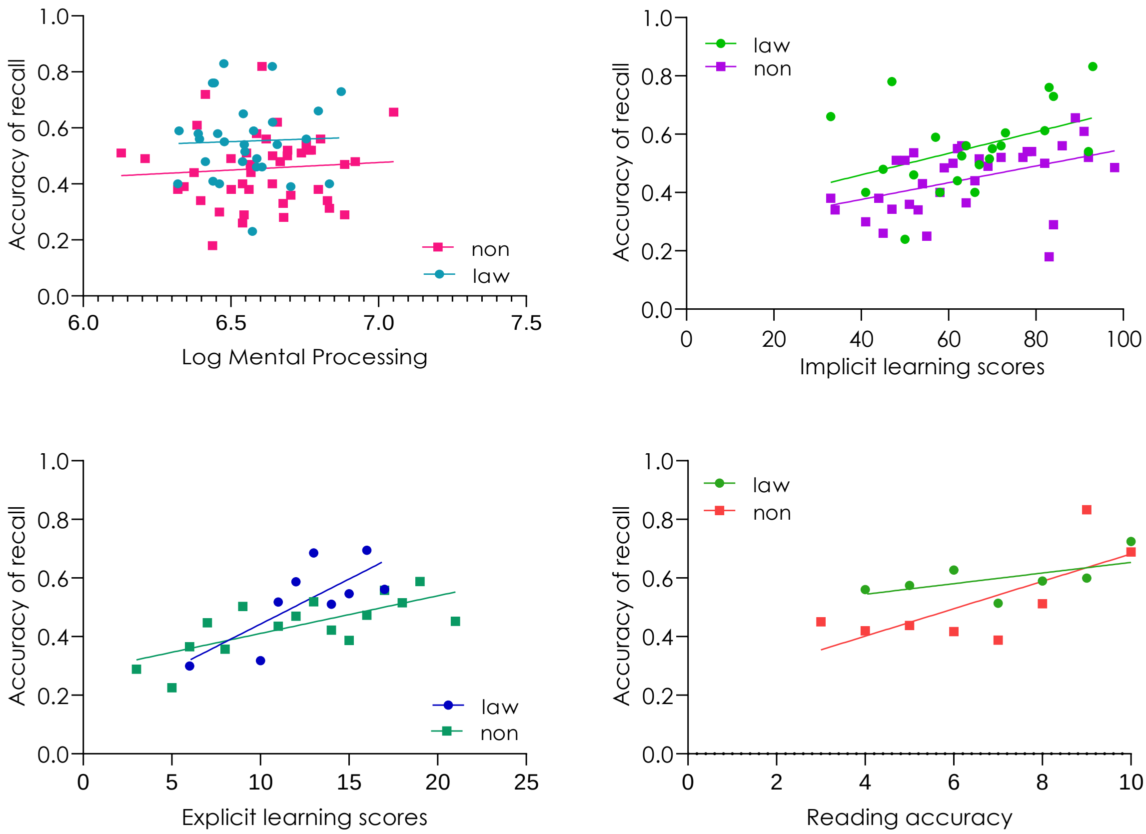


Figure 49. The effect of the cognitive factors on the accuracy of recall in Experiment IV

Figure 49 shows, unexpectedly, that mental processing speed did not significantly affect the accuracy of recall. It has to be mentioned that a different task was used to measure processing skills in experiment IV than in experiment III. Unfortunately, it is impossible to determine whether the result is due to the different method or because the processing speed, like many other factors, does not show a consistent effect. The other three factors, as expected, demonstrated a significant correlation with the accuracy of recall. The fact that both implicit and explicit learning skills were significant predictors of recall accuracy shows that some participants were better at implicit and some others better at explicit learning skills. They relied on these different skills to recall the sentences, including the target collocations[108]. Moreover, the fact that both learning skills rendered a significant effect supports the claim that individual differences play a significant role in the process of entrenchment, and we cannot fully understand it if individual differences are neglected. One of the few studies that investigated how implicit and explicit learning skills influence language skills was carried out by Llompart & Dąbrowska (2020). The study is not directly related to the one presented in this dissertation, but it includes the analysis of collocational knowledge and how it is influenced by the participants' implicit and explicit learning skills. The target items in experiments III and IV were collocations; the results obtained by Llompart & Dąbrowska (2020) can highlight which learning skills participants rely more upon for the acquisition of collocations. Their results revealed a strong and significant correlation between the two explicit memory measures, phonological short-term memory (measured by the forward Digit Span task) and memory for cross-modal associations (measured by LLAMA-B) and scores on the collocation task. These results support and, at the same time, question the results of experiment IV. A significant correlation was seen between recall accuracy and explicit learning skills. However, implicit learning skills also significantly affected the recall of the items. This difference can be accounted for by the fact that the study by Llompart & Dąbrowska (2020) investigated collocational knowledge. The current study was interested in examining which cognitive factors can facilitate the process of entrenchment, and the materials that were used to investigate the research questions were adjective-noun combinations that can be defined as collocations. For this reason, the difference between collocation learning and entrenching linguistic units could be the probable reason for the different results. Accessing learnt collocations in a test and trying to recall them relies on different processes, and these different processes might be influenced by different learning skills.

In conclusion, it can be said that the experiments' results highlighted the instability observed in the analysis of the different linguistic factors that are most often used to

[108] There is no correlation between explicit and implicit learning ($r = 0.04$), which supports the claim that some participants relied more on implicit and others more explicit learning skills.

predict language users' performance in different language-related tasks. Furthermore, the results highlighted that this instability could be due to several different reasons. First, the effect of the linguistic factors is likely to be influenced by whether the stimulus is presented in isolation or as part of a larger chunk. When words are presented in isolation, frequency is usually the winner. However, when larger linguistic units are used and embedded into sentences or texts, context-based measures such as transitional probability might show more facilitatory power than frequency. Second, the effect of these predictors largely depends on what is being tested, i.e., language acquisition, processing, or production. Some studies have revealed that participants benefited from higher contextual diversity in language-processing tasks but did not show any sensitivity towards it during word-learning tasks. These results are due to the different underlying processes that these factors influence. Transitional probabilities are likely to influence the acquisition of different patterns that emerge from usage, and frequency can facilitate learning fixed lexical expressions such as phrasal verbs or collocations. Finally, different language users might be sensitive to linguistic factors, depending on their cognitive skills. In other words, those language users who have more highly developed statistical learning skills might show more sensitivity in the learning, processing or production of words, and word combinations that occur have a high transitional probability. They might also be better at learning regular grammatical patterns. Furthermore, the interaction between the different linguistic factors showed that the different linguistic factors have a cumulative and compensatory effect when analysed together. Thus, when the word or word combination has a high frequency, the context helps predict the word, and if the language user is familiar with that item, these effects add up and do not cancel each other out. At the same time, when either frequency or contextual predictability is low, higher familiarity with the word can compensate for the lack of the frequency effect.

9.3 Finding 3

Entrenchment is operationalised in this dissertation as a phenomenon that stands on three pillars. These three pillars include the predictors of entrenchment, the processes that let language users learn, encode into memory, and strengthen the representation of linguistic items, and the product of entrenchment that shows how entrenchment affects our cognitive system. Entrenchment is a theoretical concept that can only be empirically validated if certain features or manifestations associated with it are tested and analysed. Schmid (2017a) refers to these features as the cognitive and linguistic effects and psychological affordances, and they include, *inter alia*, memory consolidation, schematisation, automatisation, chunking, emergence and increase in representational strength, faster and more effortless processing. Most of these effects are experimentally measurable. In the theoretical framework, these affordances and effects were grouped into one of the two pillars: memory consolidation, chunking, schema formation and strengthening of memory representations were seen as the processes that lead to entrenchment. Automaticity, effortless processing, and the production of words were regarded in this dissertation as being the final stage of entrenchment, i.e., its product. Two specific manifestations of automaticity were investigated across the experiments - temporal and phonetic reduction. Data on temporal and phonetic reduction was obtained from the different reading-aloud tasks.

There are several reasons why reading aloud was used to measure automaticity. Firstly, reading aloud is a complex and demanding skill. It includes the visual recognition of the word, its decoding and access from the mental lexicon and preparing its articulatory plan and the actual pronunciation of the word. All these processes have to be synchronised in order to produce fluent speech. In other words, reading aloud requires processing information in the visual and auditory modalities and relies upon word recognition skills such as phonological and orthographic decoding. Breznitz & Berman (2003) have suggested that reading aloud happens in three stages: the early perceptual, the central cognitive and the motoric stage. Automaticity can be measured in the motoric stage, but the degree of automaticity is influenced by the other two stages, i.e., how quickly participants can decode the words and access their articulatory plan. These two stages are influenced by the strength of memory representation of the words. Thus, the stronger the representation of a word in the mental lexicon is, the faster it can be accessed. Second, reading aloud includes not only language comprehension but also language production. Thus, effortless processing and production should be manifested in fluent reading.

9.3.1 The degree of automaticity is determined by the strength of memory representation.

The third main question that was asked in this dissertation is whether there is an observable connection between the strength of memory representation and speech rate. This connection would imply that the strength of memory representation should be manifested in a higher speech rate. Such a connection would validate the operationalisation of entrenchment outlined in this dissertation. It would show how different linguistic predictors lead to learning, how learning initiates memory consolidation, how subsequent exposure to the items strengthens the memory representations, and how the strength of memory representations is manifested in language processing and production.

Reading aloud was followed by a verbatim recall task in the final two experiments. There was an extra task with two main functions between the reading aloud and recall part. First, the adjectives shown to the participants after reading the sentence were possible collocates of the target noun. It was hypothesised that if participants were to select another adjective and replace the target adjective with it, the collocation would have a weaker memory representation than those collocations recalled in their original form. Second, this extra task required participants to read out the adjectives, and in this way, the capacity of the phonological loops was exceeded, which, according to Baddeley (1992; 2007), is limited to two seconds. Consequently, those recalled adjective-noun pairs should theoretically be encoded in the long-term memory as they could not be rehearsed in the short-term memory due to the intermediate task.

Combining the reading-aloud and verbatim-recall tasks allowed an investigation into the hypothesised relation between speech rate and strength of memory representation. During the recall task, the items were classified into four distinct groups: recalled, partially recalled, replaced, and not recalled. The results of the analysis of the accuracy of the recall have already shown that noun speed was a significant predictor of recall accuracy. However, the generalised mixed-effects regression model only used a binary distinction between recalled and not recalled items.

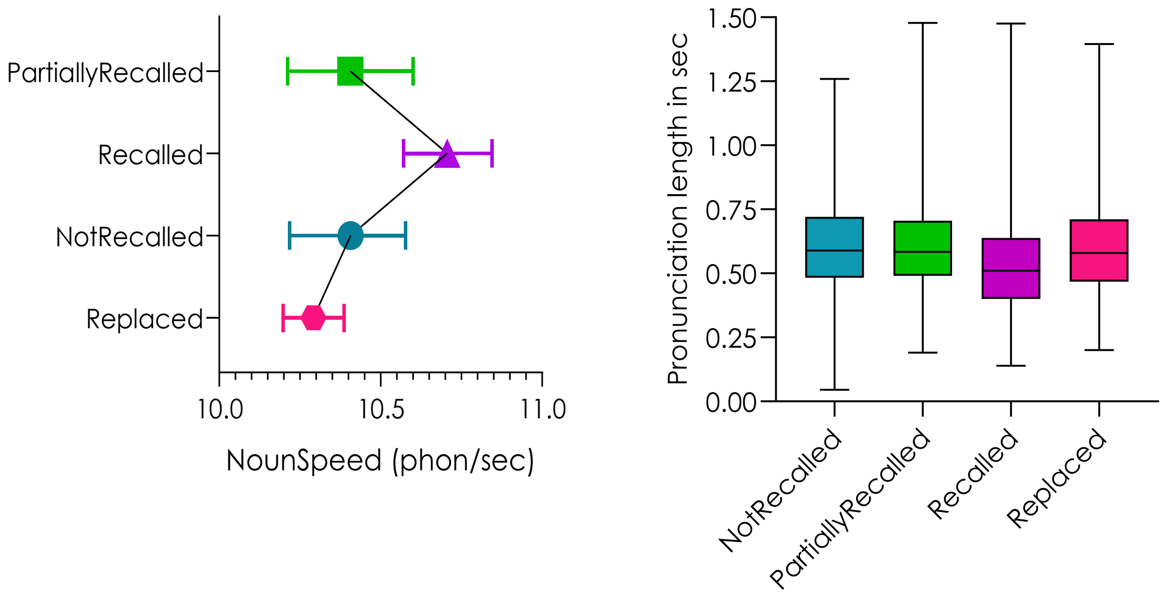


Figure 50. Correlation between the strength of memory representation and the pronunciation speed (left) and pronunciation length in seconds (right)

Figure 50 above demonstrates that words recalled in the second part of the experiment were pronounced faster during the reading-aloud task. The pronunciation rate between the recalled and not-recalled items is statistically significant ($p < 0.0001$). Moreover, it is interesting that the replaced items demonstrated the slowest pronunciation rate. This could indicate that the memory representation of the words was not strong enough and that the exposure to the additional adjectives triggered the activation of other elements, which slowed down the retrieval and pronunciation speed. It has to be emphasised that the pronunciation rate was approximately merely half a phoneme faster, but this already signifies that there should be a correlation between the two processes.

The analysis of the phonetic reduction data revealed comparable results: a positive correlation between the rate of phonetic reduction and the accuracy of memory recall. These differences are visualised in Figure 51 below.

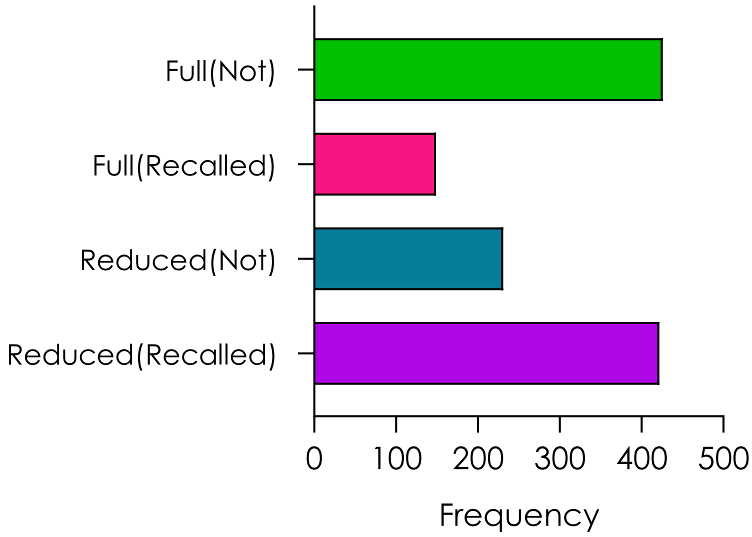


Figure 51. The relationship between phonetic reduction and strength of memory representation. Not in the parentheses means that the target words were not recalled

Words that were phonetically reduced during the reading-aloud task were more often recalled than those words that were produced in their full form. The results suggest that the temporal and phonetic reduction rate depends upon the strength of the memory representation of the words. This supports the notion that the suggested order of processes leading to entrenchment is empirically validated. First, linguistic items have to be encoded into memory and strengthened before the final stage of entrenchment, automaticity, can be achieved. The strength of memory representation depends upon the linguistic factors described above, including familiarity, frequency, and transitional probability. This consequently suggests that the three pillars of entrenchment are interconnected and form a dynamic system and that all three pillars are necessary to be able to understand how language users can extract information from the input, use it to update the existing memories and encode new memories and strengthen these representations to the degree that they can be processed and produced effortlessly. Due to the lack of research into the connection between the strength of memory representation and speech rate, only the existing theories can provide possible explanations for the results obtained. The definition of automaticity provided by Schneider & Shiffrin (1977a) emphasises that automatic processing or production is the activation of learned sequences in long-term memory. Cohen et al. (1990) and Anderson (1992) also defined automaticity as strengthening neural pathways. All these theories operationalise automatic processes as the result of strengthened memory representation. The analysis results are aligned with this operationalisation as the temporal and phonetic reduction rate strongly predicted whether the item would be recalled or not.

A further conclusion that was drawn from experiment II regarding the nature of automaticity is that it is a gradient process, but it also has an end-stage after which the process cannot become more automatic. Underlying all this is the principle that language users become better with practice while repeating the same behaviour. Once the learning has reached asymptote, the effects of the different linguistic factors wear off, and there is a progressively less noticeable gain from further practice. Thus, when automaticity, the final stage of entrenchment, is reached, frequent exposure to the items only helps to maintain this status but cannot further facilitate it, which sounds logical as there are thresholds that, due to our human nature, we cannot exceed. However, piano players can reach a speed of 20 notes per second, which is so fast that the human ear cannot perceive it anymore. This impressively fast motion is a good example of automaticity, which results from many thousands of hours of practice. However, such a speed in pronunciation would make communication unnecessarily difficult as interlocutors would have a hard time understanding each other, and, for this reason, temporal and phonetic reduction have their limits.

9.4 Interim summary

The studies presented in this section aimed to contribute to the research on entrenchment, especially how the different linguistic predictors lead to quantitative and qualitative changes in mental representations and how the strength of these mental representations is manifested in automatic processing and production. Different linguistic factors play such a crucial role in the process of entrenchment because, according to the usage-based approach, mental representations for the different linguistic items emerge from and are shaped by language usage. The linguistic factors try to capture and quantify this language usage. Some of these factors, such as frequency, focus on the pure occurrence of the words in the input, while others incorporate knowledge provided by the context, such as contextual diversity or predictability. The mental representations that emerge from language use are dynamic and change over time; frequent usage strengthens these representations, while extended periods of disuse weaken them (Langacker 1987).

It is generally accepted that language users have different experiences with language, so variation in language usage is a vital part of usage-based theories.

Nevertheless, we still know surprisingly little about how this variation is represented in the cognitive system of language users. Moreover, another open question is what are the main driving forces behind the variation observed besides the language users' experience. The language users' experience is usually quantified using different corpora. Corpora are excellent sources to describe general tendencies; however, they cannot always help investigate the variation as corpora are usually amalgamations of data from many different speakers. One of the main findings of this dissertation suggests that familiarity scores are excellent sources for investigating the degrees of entrenchment of mental representations and for gaining a more fine-grained picture of the participants' usage history.

Furthermore, the results of the experiments revealed that the effect of linguistic factors, such as frequency and transitional probability, is heavily dependent upon the processes they aim to predict, i.e., comprehension, production, and learning. Different language users also may show different sensitivity to these predictors and rely on them to a greater or lesser extent depending upon the task and the communicative situation. In addition, the results revealed that the effect of different linguistic factors could add up, demonstrating even higher facilitatory effects. Finally, the final study provided empirical validation for the theoretical framework outlined at the beginning of the dissertation, i.e., that automaticity can be viewed as the end product of entrenchment as the strength of the memory representations greatly influences whether the process is performed in a more or less automatic manner. In sum, the studies yielded here support the hypotheses that follow on from the usage-based approach, and the insights that were gained from the analysis of the interaction between the social, cognitive, and linguistic predictors are encouragements to refine the usage-based framework by studying how these factors contribute to the different processes and how the cognitive capacities of the language users shape the degree of their mental representations.

10 Conclusions

The present study has examined the relationship between linguistic, social, and cognitive factors and entrenchment on the basis of four experiments. This concluding chapter will provide a summary of this work. Firstly, the central theoretical claims regarding entrenchment will be summarised before highlighting the theoretical implication of the findings. Limitations of the current study will be discussed. Finally, this conclusion will consider recommendations and important insights for future research on entrenchment.

10.1 How can we operationalise and empirically validate entrenchment?

The central aim of chapters 3, 4 and 5 was to operationalise entrenchment, highlight how it has been defined in the previous literature, and address how a theoretical concept can be empirically validated.

Entrenchment was introduced to the linguistic community as a cognitive concept that describes the underlying process that makes human communication fluent and automatic. The frequency with which different linguistic structures occur has previously been considered because speakers and hearers can access, process, and produce language effortlessly and with remarkable speed. On this basis, with time, frequency and entrenchment became inseparable. Some definitions of entrenchment entirely rely on corpus-based frequencies and suggest a correlation between the frequency of the linguistic structure in the corpus and the degree of entrenchment (Stefanowitsch & Flach 2017). Other definitions emphasise the processes associated with entrenchment, such as chunking (De Smet & Cuyckens 2007; Blumenthal-Dramé 2012; Langacker 1987), automatisation (Langacker 1987; Bybee 2002b, 2013; Schmid 2017a; Schmid 2020; Hartsuiker & Moors 2017), strengthening of memory representations (Hilpert & Diessel 2017; Theakston 2017; Blumenthal-Dramé 2012, 2017), effortless production and processing (Bybee 2007a; Divjak & Caldwell-Harris 2015; Divjak 2019), and schematisation (Langacker 1987, 2000, 2008). However, researchers adhering to defining entrenchment through these processes also stress the significant role of frequency. Langacker (2008: 220), who introduced the concept of entrenchment, likewise underlines the crucial role frequency plays in entrenchment and proposes that linguistic units emerge via progressive entrenchment of configurations that occur in a sufficient number of communicative events to be established as cognitive routines.

Moreover, he says that regular patterns have a great advantage in terms of entrenchment and ease of activation. However, he also emphasises that a single event is enough to learn and entrench a new unit in some cases. This implies that there should be something besides frequency that leads to entrenchment. Divjak & Caldwell-Harris (2015) suggest that linguistic structures and items that have an emotional load can be entrenched after a single exposure. These studies show that entrenchment is extremely complex, including a range of linguistic predictors and cognitive processes that may not seem unrelated at first sight. The theoretical framework of this dissertation aimed to bring together the current views on entrenchment and put them into a dynamic framework that can incorporate the theoretical knowledge and the results of the empirical studies on entrenchment into one model. The resultant model is shown in Figure 52 below.

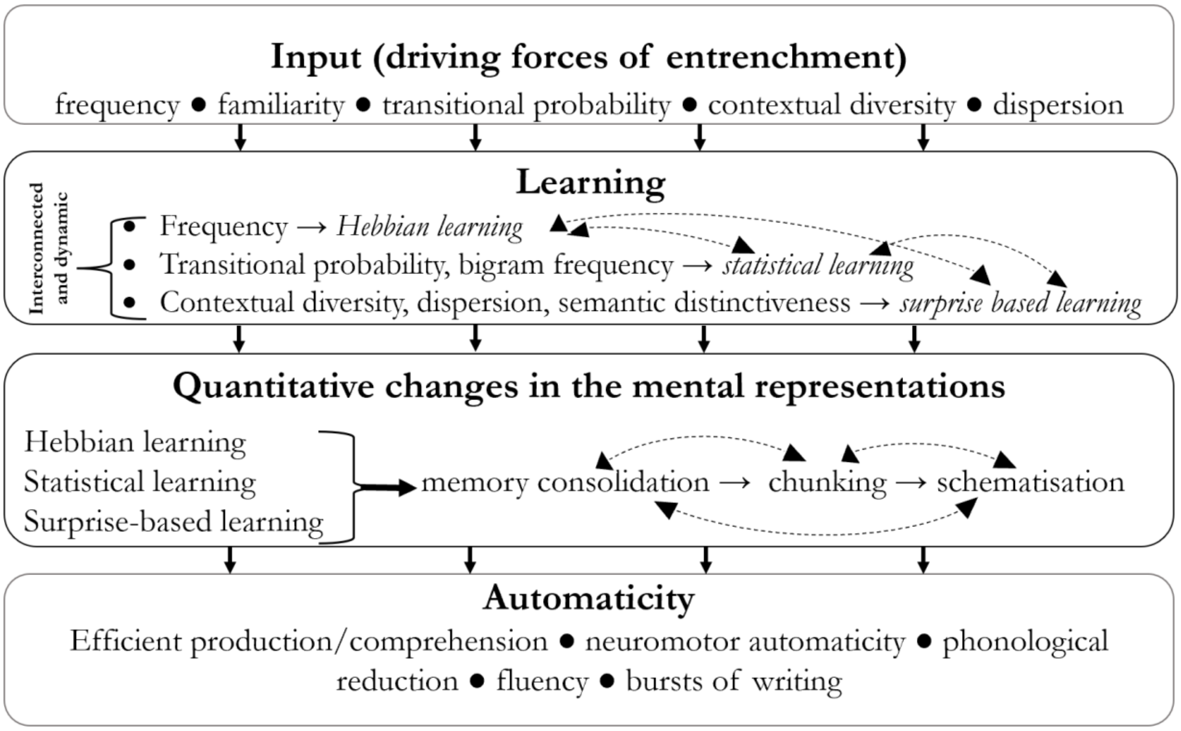


Figure 52. A dynamic model of entrenchment

The dynamic model of entrenchment suggests that besides frequency as one of the main driving forces of entrenchment, other usage-related measures contribute to the entrenchment of different linguistic structures. The input, which contains frequently occurring structures and combinations that show high-transitional probability and contextual diversity, trigger different learning processes that will enable language users to encode the new items into their memory and strengthen the representations of the old memories. Items that frequently co-occur are likely to be reorganised into chunks or schematised patterns depending on whether the item is more lexical or grammatical. Further exposure to the linguistic item or its active usage will promote strengthening, eventually leading to automatisation, which is often manifested in effortless comprehension and production. The dynamic model that is outlined here has theoretical underpinning in the usage-based approach, which emphasises that different linguistic structures emerge from usage. Speakers rely on different cognitive skills that allow them to register the occurrence of the linguistic structures, the context in which they occur, and the combinations and constructions/valency patterns they are attracted to.

Furthermore, this dynamic model emphasises the gradient nature of entrenchment. The degree of entrenchment is determined mainly by the input; however, the cognitive skills of the language users influence how attentive they are to the input. Thus, frequent exposure to the items alone is not enough for entrenchment to happen. Language users are constantly exposed to different input. Constantly processing information and encoding it into the memory is only possible if language users select a limited subset of the input. This selection mechanism is referred to as attention (Günther et al. 2017). Usually, attention is directed to salient subsets of the input. Günther et al. (2017: 289) define salience as " the conspicuity of a stimulus relative to its surrounding items". This implies that depending on the importance of the input for the current communicative situation and the language users, some items can attract more attention than others. Salience was not included in the model because it is not easily measurable, but the model does not undermine its vital role in the process of entrenchment, since it stresses the fact that there are differences among language users due to the fact that different situations can be salient for different speakers.

A further aim of the dynamic model was to establish a basis for the empirical validation of entrenchment. Interestingly,Klicken oder tippen Sie hier, um Text einzugeben. (1987) suggested that empirical validation of entrenchment is less important than its theoretical description. "While this may not be of any great importance (in the grand scheme of things), in principle [*sic*]the degree of entrenchment can be determined empirically" (Langacker 2008: 238). It is argued in this paper that the empirical validation of theoretical concepts is inevitable. Theoretical assumptions generated about language users' behaviour might largely deviate from their actual behaviour. Without empirical evidence, entrenchment, like Universal Grammar, remains a ghost, and researchers would refer to it with the questions of what it is and whether anyone has seen it[109].

Due to the numerous different phenomena associated with entrenchment, it is extremely difficult to get an experimental handle on the defining dimensions of entrenchment. This is a possible reason why not many studies specifically aim to validate the phenomenon empirically. However, as Caldwell-Harris (1997) argued, if language structures are cognitive structures that are stored in memory, then psycholinguistic techniques can be used to investigate them. All the components of the dynamic model of entrenchment are empirically testable. The four experiments that were conducted within the framework of this dissertation had the objective of testing all the components of the model. The final stage of entrenchment, automaticity, was operationalised as temporal and phonetic reduction. According to Levelt (1989), fluent speech delivery is one of humans' most automatised skills. Consequently, it provided a good basis from which to analyse the interaction between the predictors of entrenchment, the processes that contribute to it and its final manifestation. It has to be stressed that automaticity can also be manifested in written production and comprehension. Therefore, a decision was made to focus on temporal and phonetic reduction as fewer studies have investigated this phenomenon than language comprehension. The four studies can be treated as a sequence where entrenchment was conceptualised more clearly with each step, leading to a better understanding of how it can be empirically tested. Moreover, because entrenchment is subject to considerable individual differences, it was aimed in the experiments to pinpoint the source of these differences. Language users differ in their working memory capacities, learning skills, speaking styles, and attentional capacities. This suggests that different speakers can process one and the same linguistic structure differently. Some may have entrenched it as a fixed lexical expression, while others might have schematised and now use it as a pattern that has its own form and meaning. The four experiments revealed that language users not only differ in their cognitive capacities, but also to what extent they rely on the different metalinguistic data extracted from the input, such as frequency, contextual diversity, and transitional probability. Furthermore, the results of the studies showed that the driving forces of entrenchment are largely dependent on the process, that is, whether language users are learning the items, recalling, or producing them in an automatic

[109] The reference to Universal Grammar was taken from Dąbrowska (2015b)

manner. As these processes all rely on distinctive underlying cognitive mechanisms. It is suggested that the different linguistic factors foster these processes to different extents. As figure 52 suggests, frequency is likely to facilitate Hebbian learning; n-gram frequency and transitional probability in their turn promote statistical learning, and contextual diversity supports surprised based learning. However, this does not mean that they only contribute to these processes, but it is hypothesised that these measures have a more facilitatory effect. The interactions between the different linguistic factors suggest that the effect of the different linguistic factors not only adds up, but they can also compensate for each other if the linguistic structure is less frequent or, if it occurs in a combination that has a lower transitional probability, familiarity can compensate for it. The analysis of the effect of the different cognitive predictors suggest that, as hypothesised, participants differ in their cognitive skills and they rely on their most developed skills to produce and comprehend language is an effortless manner with an astonishing speed. Moreover, different language related tasks require different skills. Effortless and fluent speaking is facilitated by mental processing speed, while encoding and retrieving the linguistic structures from memory benefit from more developed implicit and explicit learning skills. Finally, the results of the experiments demonstrated that language users are subconsciously aware which linguistic structures are more and which are less entrenched. The familiarity scores assigned to the stimuli predicted their performance on the different tasks. Unlike the other linguistic predictors, familiarity showed a consistent effect across the different experiments and paradigms. The reason for this tendency is that is that familiarity scores are assigned to the different linguistic items based on their strength of memory representation, thus these are user-related measures. All the other linguistic predictors were extracted from large corpora, which usually include amalgamated data averaged over a large group of language users. These corpus-based measures provide the general tendencies observed in language usage, however, they are less likely to indicate the strength of memory representations that individual speakers have. One of the main findings of the experiment was empirical validation of the strong relation between the strength of memory representations and automaticity. Words that were recalled in experiment IV showed a tendency towards temporal and phonetic reduction. This result has empirically validated the theories on automaticity as highlighted in chapter 5. Most of these theories operationalised automaticity as the process of strengthening the representations and their fluent retrieval from memory. To the best of my knowledge, this is the first study that has investigated this correlation.

10.2 Limitations and future perspectives

Despite the many significant findings this study has produced, it also has limitations. First of all, in all the experiments, participants were enrolled or had obtained a high level of education. This implies that they all had well-developed cognitive skills, despite differences in their scores on the different cognitive tasks. To fully understand the role these cognitive skills play in the process of entrenchment in an ideal setting, participants with different social and educational backgrounds should be recruited. Second, automaticity was operationalised as temporal and phonetic reduction, and different manifestations of automaticity likely rely on different linguistic, social, and cognitive predictors. Thus, further aspects and manifestations of the same processes should be tested to develop a comprehensive model of entrenchment.

Moreover, the study focused only on processes that were hypothesised to be automatic. The process underlying the development of automatic skills should be investigated in more depth to gain a deeper understanding of automaticity. Although the aim was to incorporate most of the linguistic factors associated with entrenchment, there are further factors that contribute to the process of entrenchment that were not taken into account here, such as the age of acquisition, salience and semantic distinctiveness, among others. Finally, the studies relied on behavioural measures only. Neurolinguistic methods can further pinpoint how automatic language production and comprehension processes are executed and which mechanisms contribute to the development of this automaticity.

Based on the results and limitations of the current study, some challenges that are worth exploring in future research emerged. A fundamental challenge will be to investigate how the experiment, task, and stimuli type modulate the effect of different linguistic, social, and cognitive factors on entrenchment. As the results of the experiments and review of the previous studies showed, the source of the inconsistent effect of the different predictors is still not clear. A possible way to shed light on this inconsistency would be to test the same stimuli in different paradigms: comprehension, production and learning and compare the effects of the different predictors on these processes. Moreover, it would be interesting to investigate how entrenchment emerges and how the different linguistic and cognitive factors contribute to it at the different stages of entrenchment. Some studies have indicated that learning is facilitated to a higher degree by frequency than contextual diversity, and the opposite tendency is observed in language comprehension. A related issue is how the different modalities, i.e., visual and auditory, relate to entrenchment and whether there are individual differences between language users and their preferences towards either of these channels. It seems reasonable to assume that some speakers have more developed auditory perceptual skills and others prefer the visual domain. In other words, one of the unanswered questions is how entrenchment is modulated by modality. Visual modality allows for processing in a manner that is convenient for language users, while auditory processing has a timed nature and requires constant processing. It is hypothesised that there are differences between first and second-language learners. We learn our first language from oral exposure; however, visual stimulus likely provides more stability for second language learners due to the possible higher demand upon their cognitive resources, lack of automatised knowledge and potentially slower processing skills. A related point that could be worth investigating is how the entrenched linguistic structures facilitate the acquisition of a second language. It is hypothesised that similar constructions across the language would benefit from strong memory representations.

A last crucial issue that deserves further investigation is whether the results obtained in the four experiments would generalise to other manifestations of automaticity and whether using different stimuli, such as constructions, would lead to the same observations. The main focus of the experiments was on words and word combinations that can be placed at the lexical end of the lexico-grammatical space. The interesting question is how entrenchment manifests itself on the more grammatical end of the space. Which linguistic and cognitive factors influence the entrenchment of constructions, and how can we measure whether a construction becomes automatic? Can creativity indicate entrenchment?

In this manner, it can be said that theories of entrenchment had advanced a great deal since 1987 when first introduced to the linguistic community. We have learned more about the mechanisms that underlie entrenchment and their influential factors. Individual differences are now considered valuable sources of information rather than noise in the data. However, there is still a great deal to discover and empirically prove. Extending the analysis to an interdisciplinary perspective can potentially provide new insights into central issues of linguistics. Investigating entrenchment from the prospect of language typology, sociology, neurobiology, and anthropology can deepen our understanding of one of the main cognitive processes that allow us to acquire language.

11 References

Adams, S.G., G. Weismer & R. Kent. 1993. Speaking rate and speech movement velocity profiles. *Journal of Speech, Language, & Hearing Research* 36. 41–54.

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

Adler, Jonathan E. 2008. SURPRISE. *Educational Theory* 58(2). 149–173.

Ambridge, Ben, Evan Kidd, Caroline F. Rowland & Anna L. Theakston. 2015. The ubiquity of frequency effects in first language acquisition. *Journal of Child Language* 42(2). 239–273.

Anderson, John R. 1992. Automaticity and the ACT Theory. *The American Journal of Psychology* 105(2). 165.

Arciuli, Joanne & Ian C. Simpson. 2011. Statistical learning in typically developing children: the role of age and speed of stimulus presentation. *Developmental science* 14(3). 464–473.

Arciuli, Joanne & Ian C. Simpson. 2012. Statistical learning is related to reading ability in children and adults. *Cognitive Science* 36(2). 286–304.

Arndt, Jason. 2012. Paired-Associate Learning. In Norbert M. Seel (ed.), *Encyclopedia of the sciences of learning* (Springer reference), 2551–2552. New York: Springer.

Arnon, Inbal & Eve V. Clark. 2011. Why Brush Your Teeth Is Better Than Teeth – Children's Word Production Is Facilitated in Familiar Sentence-Frames. *Language Learning and Development* 7(2). 107–129.

Arnon, Inbal & Uriel C. Priva. 2013. More than words: the effect of multi-word frequency and constituency on phonetic duration. *Language and speech* 56(Pt 3). 349–371.

Arnon, Inbal & Neal Snider. 2010. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62(1). 67–82.

Aylett, Matthew & Alice Turk. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America* 119(5 Pt 1). 3048–3058.

Baayen, R. H., Victor Kuperman & Raymond Bertram. 2010. Frequency effects in compound processing. In Sergio Scalise & Irene Vogel (eds.), *Cross-disciplinary issues in compounding* (Amsterdam studies in the theory and history of linguistic science Series IV, Current issues in linguistic theory 311), 257–270. Philadelphia, Pa.: John Benjamins Pub. Company.

Baayen, R. Harald. 2008. *Analyzing Linguistic Data.: A Practical Introduction to Statistics using R.* New York: Cambridge University Press.

Baayen, R. Harald. 2010. Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon* 5. 436–461.

Baayen, R. Harald & Petar Milin. 2010. Analyzing Reaction Times. *International Journal of Psychological Research* 3, http://www.sfs.uni-tuebingen.de/~hbaayen/publications/BaayenMilin2010.pdf. (20 May, 2020.)

Baayen, R.H., D.J. Davidson & D.M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, http://www.sfs.uni-tuebingen.de/~hbaayen/publications/baayenDavidsonBates.pdf. (20 May, 2020.)

Baddeley, A. 1992. Working Memory: The Interface between Memory and Cognition. *Journal of Cognitive Neuroscience* 4(3).

Baddeley, Alan. 2007. *Working memory, thought and action.* (Oxford Psychology Series). New York: Oxford University Press.

Baddeley, Alan, Michael W. Eysenck & Michael C. Anderson. 2020. *Memory,* 3rd edn. London & New York: Routledge.

Baddeley, Alan D. & Graham J. Hitch. 1974. Working memory. In G.A. Bower (ed.), *The psychology of learning and motivation: Advances in research and theory*, vol. 8, 311–324. New York: Academic Press.

Baker, Chris I., Carl R. Olson & Marlene Behrmann. 2004. Role of attention and perceptual grouping in visual statistical learning. *Psychological Science* 15(7). 460–466.

Baker, Rachel E. & Ann R. Bradlow. 2009. Variability in word duration as a function of probability, speech style, and prosody. *Language and speech* 52(Pt 4). 391–413.

Bakker, Iske, Atsuko Takashima, Janet G. van Hell, Gabriele Janzen & James M. McQueen. 2015. Changes in theta and beta oscillations as signatures of novel word consolidation. *Journal of Cognitive Neuroscience* 27(7). 1286–1297.

Balota, David A. & James I. Chumbley. 1985. The Locus of Word-Frequency Effects in the Pronunciation Task: Lexical Access and/or Production? *Journal of Memory and Language* (24). 89–106.

Balota, David A., Alexander Pollatsek & Keith Rayner. 1985. The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology* 17(3). 364–390.

Balota, David A., Melvin J. Yap, Keith A. Hutchison & Michael J. Cortese. 2012. Megastudies: Large scale analysis of lexical processes. In J. S. Adelman (ed.), *Visual word recognition*, 90–115. London: Psychology Press.

Balota, D. & J. Chumbley. 1985. The locus of word-frequency effects in the pronunciation task: Lexical access and/or production?. *Journal of Memory and Language* (24.1). 89–106.

Bargh, John A. 1992. The Ecology of Automaticity: Toward Establishing the Conditions Needed to Produce Automatic Processing Effects. *The American Journal of Psychology* 105(2). 181.

Barto, Andrew, Marco Mirolli & Gianluca Baldassarre. 2013. Novelty or surprise? *Frontiers in Psychology* 4. 907.

Bates, Douglas, Martin Mächler, Ben Bolker & Walker Steve. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, https://www.jstatsoft.org/article/view/v067i01. (20 May, 2020.)

Behrens, Heike & Stefan Pfänder (eds.). 2016a. *Experience counts: frequency effects in language*. Berlin/Boston: Walter de Gruyter.

Behrens, Heike & Stefan Pfänder (eds.). 2016b. *Experience Counts: Frequency Effects in Language* (Linguae & Litterae 54). Walter de Gruyter.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009a. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1). 92–111.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009b. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1). 92–111.

Blasko, Dawn G. & Cynthia M. Connine. 1993. Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19(2). 295–308.

Blodgett, Joanna M., Rachel Cooper, Daniel H. J. Davis, Diana Kuh & Rebecca Hardy. 2021. Associations of word memory, verbal fluency, processing speed and crystallised cognitive ability with one-legged balance performance in mid and later life. *The journals of gerontology. Series A, Biological sciences and medical sciences*.
Blumenthal-Dramé, Alice. 2012. *Entrenchment in usage-based theories: What corpus data do and do not reveal about the mind.* Berlin/Boston: Walter de Gruyter.
Blumenthal-Dramé, Alice. 2017. Entrenchment from a Psycholinguistic and Neurolinguistic Perspective. In Hans Jörg Schmid (ed.), *Entrenchment and the psychology of language learning.: How we reorganize and adapt linguistic knowledge* (Language and the human lifespan), 129–153. Berlin: Walter de Gruyter.
Boersma, Paul & David Weenink. 2018. *Praat: doing phonetics by computer*.
Bogaerts, Louisa, Ram Frost & Morten H. Christiansen. 2020. Integrating statistical learning into cognitive science. *Journal of Memory and Language* 115. 104167.
Boston, Marisa Ferrara, John Hale, Reinhold Kliegl, Umesh Patil & Shravan Vasishth. 2008. Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research* 2(1). 1–12.
Bowden, Harriet W., Matthew P. Gelfand, Cristina Sanz & Michael T. Ullman. 2010. Verbal Inflectional Morphology in L1 and L2 Spanish: A Frequency Effects Study Examining Storage versus Composition. *Language Learning* 60(1). 44–87.
Brent, Michael R. & Timothy A. Cartwright. 1996. Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition* 61(1-2). 93–125.
Breznitz, Zvia & Lauren Berman. 2003. The Underlying Factors of Word Reading Rate. *Educational Psychology Review* 15, https://link.springer.com/article/10.1023/A:1024696101081.
Bridge Denckla, Marta & Laurie E. Cutting. 1999. History and significance of rapid automatized naming. *Annals of Dyslexia* 49(29). 29–42.
Brown, E. & W. Raymond. 2012. Are effects of word frequency effects of context of use? An analysis of initial fricative reduction in Spanish. *Frequency effects in language learning and processing* (1). 35–52.
Brückl, M. & W. Sendlmeier. 2003. Aging female voices: anacoustic and perceptive analysis. In C. d'Alessandroand & K. R. Scherer (eds.), *ISCA Tutorial and Research Workshop on Voice Quality: Functions, Analysis and Synthesis*.
Brysbaert, Marc & Boris New. 2009. Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41(4). 977–990.
Bunting, Michael, Nelson Cowan & John Scott Saults. 2006. How Does Running Span Work? The Quest to Measure Memory Capacity. *Quarterly Journal of Experimental Psychology* 59(10). 1691–1700.
Burgess, C. & K. Livesay. 1998. The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kucera and Francis. *Behavior Research Methods, Instruments & Computers* 30(2). 272–277.
Bybee, Joan. 2002a. Phonological Evidence for Exemplar Storage of Multiword Sequences. *SSLA* (24). 215–221.
Bybee, Joan. 2002b. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* (14). 261-290.

Bybee, Joan. 1985. *Morphology: a study of the relation between meaning and form.* Amsterdam: John Benjamins.
Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and cognitive processes* 10(5). 425–455.
Bybee, Joan. 2001. *Phonology and language use.* Cambridge: Cambridge University Press.
Bybee, Joan. 2006. From usage to grammar: the mind's response to repetition. *Language* (82(4)). 711–733.
Bybee, Joan. 2007a. *Frequency of use and the organization of language*. Oxford: Oxford University Press.
Bybee, Joan. 2007b. *Frequency of use and the organization of language*. Oxford: Oxford University Press.
Bybee, Joan. 2010. *Language, usage and cognition.* Cambridge: Cambridge University Press.
Bybee, Joan. 2013. Usage-based Theory and Exemplar Representations of Constructions. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of construction grammar*, 49–69. Oxford: Oxford Univ. Press.
Byrd, Dani. 1994. Relations of sex and dialect to reduction. *Speech Communication* 15. 39–54.
Byrd, Dani & Cheng Tan. 1996. Saying consonant clusters quickly. *Journal of Phonetics* 24. 263–282.
Caldwell-Harris, Catherine. 1997. Psycholinguistic Studies of Entrenchment. In Jean-Pierre Koenig (ed.), *Conceptual Structures, Language and Discourse*. Berkeley, CA: CSLI.
Caldwell-Harris, Catherine, Jonathan Berant & Shimon Edelman. 2012. Measuring Mental Entrenchment of Phrases with Perceptual Identification, Familiarity Ratings, and Corpus Frequency Statistics. In Dagmar Divjak & Stefan T. Gries (eds.), *Frequency effects in language representation* (Trends in linguistics244:2), 165–194. Berlin, Boston: DE GRUYTER MOUTON.
Caldwell-Harris, Catherine L. 2021. Frequency effects in reading are powerful – But is contextual diversity the more important variable? *Language and Linguistics Compass* 15(12).
Caramazza, Alfonso, Albert Costa, Michele Miozzo & Yanchao Bi. 2001. The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* (27.14). 30–50.
Cattell, James M. 1886. The time it takes to see and name objects. *Mind* 41. 63–65.
Chang, Franklin, Evan Kidd & Caroline F. Rowland. 2013. Prediction in processing is a by-product of language learning: [Commentary on Pickering & Garrod: An integrated theory of language production and comprehension]. *Behavioral and Brain Sciences* 36(4). 350–351.
Chenoweth, Ann N. & John R. Hayes. 2001. Fluency in Writing. *Written Communication* 18(1). 80–98.
Cholin, Joana, Gary S. Dell & Willem J. M. Levelt. 2011. Planning and articulation in incremental word production: syllable-frequency effects in English. *Journal of experimental psychology. Learning, memory, and cognition* 37(1). 109–122.
Christiansen, Morten H. & Nick Chater. 2016. The Now-or-Never bottleneck: A fundamental constraint on language. *The Behavioral and brain sciences* 39. e62.

Clopper, Cynthia G. & Rory Turnbull. 2018. Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In Francesco Cangemi, Meghan Clayards & Oliver Niebuhr (eds.), *Rethinking reduction*, 25–72. Berlin: DE GRUYTER MOUTON.
Coetzee, Andries W. & Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language & Linguistic Theory* 31, https://link.springer.com/article/10.1007/s11049-012-9179-z.
Cohen, J. D., K. Dunbar & J. L. McClelland. 1990. On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychological Review* 97(3). 332–361.
Conklin, Kathy & Norbert Schmitt. 2012. The Processing of Formulaic Language. *Annual Review of Applied Linguistics* 32. 45–61.
Connine, Cynthia M., John Mullennix,, Eve Shernoff & Jennifer Yelen. 1990. Word familiarity and frequency in visual and auditory lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16(6). 1084–1096.
Conwey, Andrew R.A., Michael J. Kane, Michael F. Bunting, Zach D. Hamrick & Oliver Wilhelm. 2005. Working memory span tasks: A methodological review and users guide. *Psychonomic Bulletin & Review* 12(5). 769–786.
Cronk, Brian C., Susan D. Lima & Wendy A. Schweigert. 1993. Idioms in sentences: Effects of frequency, literalness, and familiarity. *Journal of Psycholinguistic Research* 22(1). 59–82.
Curl, Traci S., John Local & Gareth Walker. 2006. Repetition and the prosody–pragmatics interface. *Journal of Pragmatics* 38(10). 1721–1751.
Dąbrowska, Ewa. 2012. Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism* 2. 219–253.
Dąbrowska, Ewa. 2015a. Individual Differences in Grammatical Knowledge. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*, 650–668. Berlin: DE GRUYTER MOUTON.
Dąbrowska, Ewa. 2015b. What exactly is Universal Grammar, and has anyone seen it? *Frontiers in Psychology* 6. 852.
Dąbrowska, Ewa. 2015c. Words that Go Together. *The Mental Lexicon* 9(3). 401–418.
Dalby, Jonathan M. 1986. *Phonetic Structure of Fast Speech in American English.* Indiana: Bloomington.
Daneman, Meredyth & Patricia A. Carpenter. 1980. Individual Differences in Working Memory and Reading. *Journal of Verbal Learning and Verbal Behavior* 19. 450–466.
Dankovičová, J. 1997. The domain of articulation rate variation. *Journal of Phonetics* 25. 287–312.
Davidson, Lisa. 2006. Schwa Elision in Fast Speech: Segmental Deletion or Gestural Overlap. *Phonetica* 63. 79–112.
de Jong, Nivja & Ton G. Wempe. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods* 41(2). 385–390.
De Jong, Nivja H. & Hans Rutger Bosker. 2013. Choosing a threshold for silent pauses to meaure second language fluency.
de Pijper, Jan-Roelof & Angelien A. Sanderman. 1994. On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America* 96. 2037–2047.
De Smet, Hendrik. 2017. Entrenchent Effects in Language Change. In Hans Jörg Schmid (ed.), *Entrenchment and the psychology of language learning.: How we reorganize*

and adapt linguisitc knowledge (Language and the human lifespan), 75–101. Berlin: Walter de Gruyter.
De Smet, Hendrik & Hubert Cuyckens. 2007. Diachronic aspects of complementation_ Constructions, entrenchment, and the matching problem. In C.M. Cain & G. Russom (eds.), *Studies in the history Studies in the history of the English language III: Managing chaos: Strategies for identifying change in English*, 187–214. DE GRUYTER MOUTON.
Dehaene, Stanislas & Jean-Pierre Changeux. 2004. Neural mechanisms for access to consciousness. In Michael S. Gazzaniga (ed.), *The cognitive neurosciences*, 3rd edn., 1145–1158. Cambridge, Mass., London: MIT Press.
Dehaene, Stanislas, Florent Meyniel, Catherine Wacongne, Liping Wang & Christophe pallier. 2015. The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron* 88, https://www.sciencedirect.com/science/article/pii/S089662731500776X.
Dell, G. 1990. Effects of frequency and vocabulary type on phonological speech errors. *Language and cognitive processes* (5.4). 313–349.
Demberg, Vera & Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2). 193–210.
Diessel, Holger. 2019. *The Grammar Network.* Cambridge University Press.
Divjak, Dagmar. 2019. *Frequency in Language.* New York: Cambridge University Press.
Divjak, Dagmar & Catherine L. Caldwell-Harris. 2015. Frequency and entrenchment. In Ewa Da˛browska & Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics* (Handbücher zur Sprach- und Kommunikationswissenschaft 39), 53–75. Berlin/Boston: DE GRUYTER MOUTON.
Divjak, Dagmar & Stefan T. Gries (eds.). 2012. *Frequency effects in language representation* (Trends in linguistics244:2). Berlin, Boston: DE GRUYTER MOUTON.
Editorial. 2011. An introduction to reduced pronunciation variants. *Journal of Phonetics* 39(3). 253–260.
Ehrlich, Susan F. & Keith Rayner. 1981. Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior* 20(6). 641–655.
Ellis, Nick. 2002. Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* (24). 143–188.
Ellis, Nick. 2013. Second language acquisition. In Trousdale G. & T. Hoffmann (ed.), *Oxford Handbook of Construction Grammar*, 365–378. Oxford: Oxford University Press.
Ellis, Nick. 2016. Frequency in language learning and language change. In Heike Behrens & Stefan Pfänder (eds.), *Experience counts: frequency effects in language*, 239–256. Berlin/Boston: Walter de Gruyter.
Ellis, Nick C. 1996. SEQUENCING IN SLA: Phonological Memory, Chunking, and Points of Order. *Studies in Second Language Acquisition* 18, http://www.jstor.org/stable/44487860.
Ellis, Nick C. 2003. Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure. In Catherine J. Doughty & Michael H. Long (eds.), *The Handbook of Second Language Acquisition*, 63–103. Oxford, UK: Blackwell Publishing Ltd.
Ellis, Nick C. 2008. Implicit and explicit knowledge about language. *Encyclopedia of language and education* 6. 1–13.

Ellis, Nick C. 2015. *Implicit and explicit learning: Their dynamic interface and complexity*. John Benjamins.
Ellis, Nick C. 2017. Chunking in Language Usage, Learning and Change: I Don't Know. In Marianne Hundt, Sandra Mollin & Simone E. Pfenninger (eds.), *The changing English language: Psycholinguistic perspectives* (Studies in English language), 113–147. Cambridge: Cambridge University Press.
Ellis, Nick C., Matthew B. O'Donnell & Ute Römer. 2014. The processing of verb-argument constructions is sensitive to form, function, frequency, contingency and prototypicality. *Cognitive Linguistics* 25(1). 55–98.
Endress, Ansgar D. & Alan Langus. 2017. Transitional probabilities count more than frequency, but might not be used for memorization. *Cognitive Psychology* 92. 37–64.
Ernestus, Mirjam & Natasha Warner. 2011. An introduction to reduced pronunciation variants. *Journal of Phonetics* 39. 253–260.
Evert, Stefan. 2009. Corpora and collocations. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, vol. 2, 1212–1248. Berlin, New York: Mouton de Gruyter.
Evert, Stefan & Brigitte Krenn. 2001. Methods for the Qualitative Evaluation of Lexical Association Measures. 188–195.
Fischer, Volker & Siegfried Kunzmann. 2013. The EML Transcription Platform – a flexible transcription environment for robust speech recognition. 2060–2066.
Fiser, József & Richard N. Aslin. 2002. Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences* 99(24). 15822–15826.
Flemming, Edward. 2009. The phonetics of schwa vowels. In C. Jones & Donka Minkova (eds.), *Phonological Weakness in English: From Old to Present-Day English*, 78–98. London: Palgrave Macmillan.
Foster, Meadhbh I. & Mark T. Keane. 2019. The Role of Surprise in Learning: Different Surprising Outcomes Affect Memorability Differentially. *Topics in cognitive science* 11(1). 75–87.
Frank, Stefan L. & Rens Bod. 2011. Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science* 22(6). 829–834.
Frost, Ram, Blair C. Armstrong, Noam Siegelman & Morten H. Christiansen. 2015. Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences* 19(3). 117–125.
Gahl, Susanne. 2008. Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84(3). 474–496.
Gahl, Susanne & Susan M. Garnsey. 2004. Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*. 748–775.
Gahl, Susanne, Yao Yao & Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66(4). 789–806.
Garagnani, Max, Thomas Wennekers & Friedemann Pulvermüller. 2008. A neuroanatomically grounded Hebbian-learning model of attention-language interactions in the human brain. *The European journal of neuroscience* 27(2). 492–513.

Gardner, Michael K., E.Z. Rothkopf, Richard Lapan & Toby Lafferty. 1987. The word frequency effect in lexical decision: Finding a frequency-based component. *Memory & Cognition* 15(1). 24–28.

Garibyan, Armine. *Processing collocations cross-linguistically: a psycholinguistic and neurolinguistic study*.

Gathercole, Susan E. & Alan D. Baddeley. 1993. *Working memory and Language* (Essay in Congnitive Psychology). New York: Psychology Press.

Gernsbacher, M. A. 1984. Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness and polysemy. *Journal of Experimental Psychology: General* 113. 256–281.

Gilchrist, Amanda L. 2015. How should we measure chunks? a continuing issue in chunking research and a way forward. *Frontiers in psychology* 6.

Gilquin, Gaëtanelle. 2020. Belgian Journal of Linguistics, Volume 34 (2020). *Belgian Journal of Linguistics* 34. 99–109.

Glenberg, A. M. 1979. Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition* 7, https://link.springer.com/article/10.3758/BF03197590.

Glenberg, Arthur M. 1976. Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior* 15(1). 1–16.

GOBET, F., P. LANE, S. CROKER, P. CHENG, G. JONES, I. OLIVER & J. PINE. 2001. Chunking mechanisms in human learning. *Trends in Cognitive Sciences* 5(6). 236–243.

Godfrey, John J., Edward C. Holliman & Jane McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. *IEEE* 1. 517–520.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences* 7(5). 219–224.

Goldberg, Adele E. 2006. *Constructions at Work: the nature of generalization in Language*. Oxford: Oxford University Press.

Goldberg, Adele E. 2019. *Explain me this: Creativity, Competition and the Partial Productivity of Constructions*. Princeton: Princeton University Press.

Goujon, Annabelle & Joel Fagot. 2013. "Learning of spatial statistics in nonhuman primates: contextual cueing in baboons (Papio papio). *Behavioural Brain Research* 247. 101–109.

Gradoville, Michael. 2017. The cognitive representation of multi-word sequences: A usage-based approach to the reduction of Fortalezense Portuguese para. *Lingua* 199. 94–116.

Grantham O'Brian, Mary & Sarah M. B. Fagan. 2016. *German Phonetics and Phonology: Theory and Practise*. New Haven & London: Yale University Press.

Greenwald, A. G. 1992. New look 3. Unconscious cognition reclaimed. *The American psychologist* 47(6). 766–779.

Gregory, Michelle L., William D. Raymond, Alan Bell, Eric Fosler-Lussier, & Daniel Jurafsky. 1999. The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society* 35. 151–166.

Gries, Stefan T. 2008. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13(4). 403–437.

Grosjean, F. & M. Collins. 1979. Breathing, pausing and reading. *Phonetica* 36(2). 98–114.
Günther, Franziska, Hermann J. Müller & Thomas Geyer. 2017. Salience, attention, and perception. In Hans-Jörg Schmid (ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge / edited by Hans-Jörg Schmid* (Language and the human lifespan), 289–312. Berlin: DE GRUYTER MOUTON.
Guy, Gregory R. 1992. Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change* 3. 1–22.
Haber, Lyn R. & Ralph Norman Haber. 1982. Does Silent Reading Involve Articulation? Evidence from Tongue Twisters. *The American Journal of Psychology* 95(3). 409–419.
Hartsuiker, Robert J. & Agnes Moors. 2017. On the Automaticity of Language Processing. In Hans Jörg Schmid (ed.), *Entrenchment and the psychology of language learning.: How we reorganize and adapt linguistic knowledge* (Language and the human lifespan), 201–227. Berlin: Walter de Gruyter.
Heinz, Wolfgang. 2004. Das deutsche Strafverfahren–rechtliche Grundlagen, rechtstatsächliche Befunde, historische Entwicklung und aktuelle Tendenzen, http://www. uni-konstanz. de/rtf/kis/strafverfahren. htm. (2 February, 2019.)
Herbst, Thomas. 2010. *English Linguistics: A Coursebook for Students of English.* Berlin: DE GRUYTER MOUTON.
Herbst, Thomas & Susen Schüller. 2008. *Introduction to Syntactic Analysis: A Valancy Approach.* Tübingen: Gunter Narr Verlag Tübingen.
Hills, Thomas T., Josita Maouene, Brian Riordan & Linda B. Smith. 2010. The Associative Structure of Language: Contextual Diversity in Early Word Learning. *Journal of Memory and Language* 63(3). 259–273.
Hilpert, Martin & Holger Diessel. 2017. Entrenchment in Construction Grammar. In Hans Jörg Schmid (ed.), *Entrenchment and the psychology of language learning.: How we reorganize and adapt linguistic knowledge* (Language and the human lifespan), 57–75. Berlin: Walter de Gruyter.
Hochberg, J. 1976. Toward a speech-plan eye-movement model of reading. In R. A. Monty & J. W. Senders (eds.), *Eye movements and psychological processes*, 397–416. New Jersey: Erlbaum.
Holdnack, James A., Aurelio Prifitera, Lawrence G. Weiss & Donald H. Saklofske. 2015. WISC-V and the Personalized Assessment Approach. In James A. Holdnack, Aurelio Prifitera, Lawrence G. Weiss & Donald H. Saklofske (eds.), *WISC-V Assessment and Interpretation: Scientist-Practitioner Perspectives*, 373–413. New Jersey: Academic Press.
Hudson, Richard A. 1996. *Sociolinguistics.* Cambridge: Cambridge University Press.
Huettig, Falk. 2015. Four central questions about prediction in language processing. *Brain Research* 1626. 118–135.
Huettig, Falk & Nivedita Mani. 2016. Is prediction necessary to understand language? Probably not. *LANGUAGE, COGNITION AND NEUROSCIENCE* 31(1). 19–31.
Hunt, Ruskin H. & Richard N. Aslin. 2001. Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General* 130(4). 658–680.
Jacewicz, Ewa, Robert A. Fox, Caitlin O'Neill & Joseph Salmons. 2010. Articulation rate across dialect, age, and gender. *Lang Var Change* 21(2). 233–256.

Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press.
Jaeger, T. Florian & Kodi Weatherholtz. 2016. What the Heck Is Salience? How Predictive Language Processing Contributes to Sociolinguistic Perception. *Frontiers in Psychology* 7. 1115.
Jakobson, Roman & Morris Halle. 1956. *Fundamentals of language.'s-Gravenhage: Mouton*. The Hague: Mouton & Co.
Jande, P. A. 2003. Evaluating rules for phonological reduction in Swedish. *PHONUM* 9. 149–152.
Jegerski, Jill & Bill VanPatten. 2013. *Research methods in second language psycholinguistics*. New York / London: Routledge.
Jescheniak, Jörg D. & Willem J. M. Levelt. 1994. Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of experimental psychology. Learning, memory, and cognition* 20(4). 824–843.
Jiang, Nan. 2012. *Conducting reaction time research in second language studies*. New York: Routledge.
Johns, Brendan T., Thomas M. Gruenenfelder, David B. Pisoni & Michael N. Jones. 2012. Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *The Journal of the Acoustical Society of America* 132(2). 74-80.
Johnson, K. 2004. Massive reduction in conversational American English. 29–54.
Jones, Michael N., Brendan T. Johns & Gabriel Recchia. 2012. The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 66(2). 115–124.
Jost, Ethan & Christiansen H. Morten. 2017. Statistical Learning as a Domain-General Mechanism of Entrenchment. In Hans Jörg Schmid (ed.), *Entrenchment and the psychology of language learning.: How we reorganize and adapt linguistic knowledge* (Language and the human lifespan), 227–245. Berlin: Walter de Gruyter.
Juhasz, Barbara J., Yun-Hsuan Lai & Michelle L. Woodcock. 2015. A database of 629 English compound words: ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods* 47(4). 1004–1019.
Juhasz, Barbara J. & Keith Rayner. 2003. Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29(6). 1312–1318.
Jurafsky, Daniel. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science* 20(2). 137–194.
Jurafsky, Daniel. 2003. Probabilistic modelling in psycholinguistics: Linguistic comprehension and production. *Probabilistic linguistics* 21, https://web.stanford.edu/~jurafsky/prob.pdf. (February 2021.)
Jurafsky, Daniel, Alan Bell, Michelle Gregory & William Raymond. 2001a. Probabilistic relations between words: Evidence from reduction in lexical production. *Typological studies in language* (45). 229–254.
Jurafsky, Daniel, Alan Bell, Michelle Gregory & William D. Raymond. 2001b. The effect of language model probability on pronunciation reduction. *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference* 2.

Just, Marcel A., Patricia A. Carpenter & Jacqueline D. Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111(2). 228–238.

Kahn, Jason M. & Jennifer E. Arnold. 2015. Articulatory and lexical repetition effects on durational reduction: speaker experience vs. common ground. *LANGUAGE, COGNITION AND NEUROSCIENCE* 30(1-2). 103–119.

Karinne Sauval & Fabienne Chetail. 2017. *Does bigram frequency genuinely impact visual word recognition? An ERP study*.

Kathryn Bock, Gary S. Dell, Susan M. Garnsey, Arthur F. Kramer & Tate T. Kubose. 2007. Car talk, car listen. In *Automaticity and Control in Language Processing*, 39–60. Psychology Press.

Kemper, Susan, Ruth E. Herman & Cindy H. Lian. 2003. The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech of noise. *Psychology and aging* 18(2). 181–192.

Kestens, Katrien, Sofie Degeest, Marijke Miatton & Hannah Keppler. 2021. Visual and Verbal Working Memory and Processing Speed Across the Adult Lifespan: The Effect of Age, Sex, Educational Level, Awakeness, and Hearing Sensitivity. *Frontiers in Psychology* 12, https://www.frontiersin.org/articles/10.3389/fpsyg.2021.668828/full.

Kilgarriff, Adam, Baisa Vít, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: ten years on. *Lexicography* 1. 7–36.

Kleinschmidt, Dave F. & T. Florian Jaeger. 2015. Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review* 122(2). 148–203.

Kohler, K. J. 1977. *Einführung in die Phonetic des Deutschen* (Grundlagen der Germanistik 20). Neuburg: Erich Schmidt Verlag.

Kohler, K. J. 1991. The Phonetics/Phonology Issue in the Study of Articulatory Reduction. *Phonetica* 48((2-4)). 180–192.

Kohler, K.J. 1990. Segmental Reduction in Connected Speech in German: Phonological Facts and Phonetic Explanations. In Hardcastle W.J. & Marchal A. (eds.), *Speech production and speech modelling* (Behavioural and social sciences 55), 69–92. Dordrecht, Netherlands, Boston: Kluwer Academic Publishers.

Kouider, Sid & Stanislas Dehaene. 2007. Levels of processing during non-conscious perception: a critical review of visual masking. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 362(1481). 857–875.

Kuperberg, Gina R. & T. Florian Jaeger. 2015. What do we mean by prediction in language comprehension? *LANGUAGE, COGNITION AND NEUROSCIENCE*, http://dx.doi.org/10.1080/23273798.2015.1102299. (28 April, 2020.)

Kutas, Marta & Kara D. Federmeier. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology* 62. 621–647.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82, https://www.jstatsoft.org/article/view/v082i13/0. (20 May, 2020.)

Labov, William. 1972. *Language in the inner city; studies in the Black English vernacular: Studies in the Black English vernacular* (Conduct and communication 3). Philadelphia: Univ. of Pennsylvania Press.

Lane, Liane W., Michelle Groisman & Victor S. Ferreira. 2006. Don't talk about pink elephants! Speaker's control over leaking private information during language production. *Psychological Science* 17(4). 273–277.

Langacker, Ronald. 1987. *Foundations of cognitive grammar: Theoretical prerequisites.* Stanford: Stanford University Press.

Langacker, Ronald. 2000. A dynamic usage-based model. *Usage-based models of language*. 1–63.

Langacker, Ronald. 2008. *Cognitive Grammar: A basic introduction.* Oxford: Oxford University Press.

Leijten, Mariëlle & Luuk van Waes. 2013. Keystroke Logging in Writing Research. *Written Communication* 30(3). 358–392.

Levelt, W. J. M. 1989. *Speaking: From intention to articulation / Willem J.M. Levelt* (ACL-MIT Press series in natural-language processing). Cambridge, Mass.: MIT Press.

Levshina, Natalia (ed.). 2015. *How to do Linguistics with R.* Amsterdam: John Benjamins Publishing Company.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. *Speech production and speech modelling.* 403–439.

Linzen, Tal & T. Florian Jaeger. 2015. Uncertainty and Expectation in Sentence Processing: Evidence From Subcategorization Distributions. *Cognitive Science*, https://www.researchgate.net/publication/281146046_Uncertainty_and_Expectation_in_Sentence_Processing_Evidence_From_Subcategorization_Distributions. (20 May, 2020.)

Llompart, Miquel & Ewa Dąbrowska. 2020. Explicit but Not Implicit Memory Predicts Ultimate Attainment in the Native Language. *Frontiers in Psychology* 11. 569586.

Lorenz, David & David Tizón-Couto. 2019. Chunking or predicting – frequency information and reduction in the perception of multi-word sequences. *Cognitive Linguistics* 30, https://doi.org/10.1515/cog-2017-0138. (28 April, 2020.)

MacDonald, Maryellen C. 1993. The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language* 32(5). 692–715.

Machelett, Kirsten. 1996. Das Lesen von Sonagrammen V1.0: Begleitendes Hypertext-Dokument zur Vorlesung. https://www.phonetik.uni-muenchen.de/studium/skripten/SGL/SGLHomeBody.html.

Mak, Matthew H., Yaling Hsiao & Kate Nation. 2021. Anchoring and contextual variation in the early stages of incidental word learning during reading. *Journal of Memory and Language* 118. 1–24.

McCauley, Stewart M. & Morten H. Christiansen. 2015. Individual Differences in Chunking Ability Predict On-line Sentence Processing. https://www.researchgate.net/profile/Stewart-Mccauley/publication/275952546_Individual_differences_in_chunking_ability_predict_on-line_sentence_processing/links/554a7d020cf29f836c965141/Individual-differences-in-chunking-ability-predict-on-line-sentence-processing.pdf. (5 March, 2021.)

McConnell, Kyla & Alice Blumenthal-Dramé. 2019. Effects of task and corpus-derived association scores on the online processing of collocations. *Corpus Linguistics and Linguistic Theory*, https://doi.org/10.1515/cllt-2018-0030. (28 April, 2020.)

McConnell, Kyla & Alice Blumenthal-Dramé. 2021. Usage-Based Individual Differences in the Probabilistic Processing of Multi-Word Sequences. *Frontiers in Communication* 6.

McDonald, Scott A. & Richard C. Shillcock. 2003. Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science* 14(6). 648–652.

McDowd, Joan, Tze-Chun Tang, Pei-Chun Tsai, Shing-Yaw Wang & Chwen-Yng Su. 2011. The association between verbal memory, processing speed, negative symptoms and functional capacity in schizophrenia. *Psychiatry research* 187, https://pubmed.ncbi.nlm.nih.gov/21320727/.

McRae, Ken, Michael J. Spivey-Knowlton & Michael K. Tanenhaus. 1998. Modelling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language* 38(3). 2283–2312.

Miller, George A. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review* 63(81-97).

Misyak, Jennifer B. & Morten H. Christiansen. 2012. Statistical Learning and Language: An Individual Differences Study. *Language Learning* 62(1). 302–331.

Moers, Cornelia, Antje Meyer & Esther Janse. 2017. Effects of Word Frequency and Transitional Probability on Word Reading Durations of Younger and Older Speakers. *Language and speech* 60(2). 289–317.

Monsell, Stephen. 2003. Task switching. *Trends in Cognitive Sciences* 7(3). 134–140.

Moors, Agnes. 2016. Automaticity: Componential, Causal, and Mechanistic Explanations. *Annual review of psychology* 67. 263–287.

Mooshammer, Christine & Christian Geng. 2008. Acoustic and articulatory manifestations of vowel reduction in German. *Journal of the International Phonetic Association* 38(2). 117–136.

Mousikou, Petroula & Kathleen Rastle. 2015. Lexical frequency effects on articulation: a comparion of picture naming and reading aloud. *Frontiers in psychology* 6, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4606046/. (11 October, 2019.)

MRC Psycholinguistic Database. 2006. http://www.psy.uwa.edu.au/mrcdata. (February 2021.)

Mukherjee, Joybrato. 2005. *English ditransitive verbs: Aspects of theory, description and a usage-based model.* Amsterdam: Rodopi.

Munnich, Edward, Michael A. Ranney & Mirian Song. 2007. Surprise, Surprise: The Role of Surprising Numerical Feedback in Belief Change. 503–508.

Munson, Benjamin. 2007. Lexical characteristics mediate the influence of sex and sex typicality on vowel-space size. 885–888.

Munson, Benjamin & Nancy P. Solomon. 2004. The Effect of Phonological Neighborhood Density on Vowel Articulation. *Journal of Speech, Language, and Hearing Research* 47(5). 1048–1058.

Myers, James & Yingshing Li. 2009. Lexical frequency effects in Taiwan Southern Min syllable contraction. *Journal of Phonetics* 37(2). 212–230.

Narayanan, Srini & Daniel Jurafsky (eds.). 1998. *Bayesian models of human sentence processing*.

New, Boris, Ludovic Ferrand, Christophe pallier & Marc Brysbaert. 2006. Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review* 13, https://link.springer.com/article/10.3758/BF03193811. (20 May, 2020.)

Newell, Allen. 1990. *Unified theories of cognition.* Cambridge, MA: MIT Press.

Nusbaum, Howard & Marina Dedina. 1985. The effects of word frequency and subjective familiarity on visual lexical decisions. *Research on Speech Perception, Progress Resport* 11. 1084–1096.

O'Brien, Mary G. & Sarah M. B. Fagan. 2016. *German phonetics and phonology: Theory and practice / Mary Grantham O'Brien, Sarah M.B. Fagan.* New Haven: Yale University Press.

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

Papagno, Costanza, Tim Valentine & Alan Baddeley. 1991. Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language* 30(3). 331–347.

Parrell, Benjamin, Adam C. Lammert, Gregory Ciccarelli & Thomas F. Quatieri. 2019. Current models of speech motor control: A control-theoretic overview of architectures and properties. *The Journal of the Acoustical Society of America* 145(3). 1456.

Patterson, D., P.C. LoCasta & C.M. Connine. 2003. Corpora analyses of frequency of schwa delition in conversational American English. *Phonetica* 60. 45–69.

Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman & Jonas Kristoffer Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51(1). 195–203.

Pelucchi, Bruna, Jessica F. Hay & Jenny R. Saffran. 2009a. Learning in reverse: eight-month-old infants track backward transitional probabilities. *Cognition* 113(2). 244–247.

Pelucchi, Bruna, Jessica F. Hay & Jenny R. Saffran. 2009b. Statistical learning in a natural language by 8-month-old infants. *Child Development* 80(3). 674–685.

Perea, Manuel, Ana P. Soares & Montserrat Comesaña. 2013. Contextual diversity is a main determinant of word identification times in young readers. *Journal of experimental child psychology* 116(1). 37–44.

Perruchet, Pierre & Stéphane Desaulty. 2008. A role for backward transitional probabilities in word segmentation? *Memory & Cognition* 36(7). 1299–1305.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In Joan Bybee & Paul J. Hopper (eds.), *Frequency effects and the emergence of lexical structure*. Amsterdam: John Benjamins.

Pierrehumbert, Janet. 2002. Word-specific phonetics. In Gussenhoven C. & N. Warner (eds.), *Laboratory Phonology 7* (Phonology and Phonetics), 101–139. Berlin: DE GRUYTER MOUTON.

Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005. Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America* 118(4). 2561–2569.

Polišenská, Kamila, Shula Chiat & Penny Roy. 2015. Sentence repetition: what does the task measure? *International journal of language & communication disorders* 50(1). 106–118.

Pulvermüller, Friedemann. 1999. Words in the brain's language. *Behavioral and Brain Sciences* 22. 253–336.

Raymond, W., R. Dautricourt & E. Hume. 2006. Word-internal/t, d/deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change* (18.1). 55–97.

Rayner, Keith. 2006. Vowel Processing During Silent Reading: Evidence From Eye Movements. *Journal of Experimental Psychology: Copyright 2006 by the American Psychological Association Learning, Memory, and Cognition* 32(2). 416–424.
Rayner, Keith, Katherine S. Binder, Jane Ashby & Alexander Pollatsek. 2001. Eye movement control in reading: Word predictability has little influence on initial landing positions in words. *Vision Research* 41(7). 943–954.
Reber, Arthur S. & Rhianon Allen. 2013. Individual Differences in Implicit Learning: Implications for the Evolution of Consciousness. In Steven M. Miller (ed.), *The Constitution of Visual Consciousness* (Advances in Consciousness Research 90), vol. 20, 227. Amsterdam: John Benjamins Publishing Company.
Reber, Paul J. 2013. The neural basis of implicit learning and memory: a review of neuropsychological and neuroimaging research. *Neuropsychologia* 51, https://www.sciencedirect.com/science/article/pii/S002839321300208X.
Rescorla, Robert A. & Allan R. Wagner. 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black & W.F. Prokasy (eds.), *Classical Conditioning II*, 64–99. New York: Appleton-Century-Crofts.
Robinson, Peter. 1995. Attention, Memory, and the “Noticing” Hypothesis. *Language Learning* 45(2). 283–331.
Rogers, Timothy T. & James L. McClelland. 2004. *Semantic cognition: A parallel distributed processing approach*. Cambridge: MIT Press.
Roland, Douglas & Daniel Jurafsky. 2000. Verb sense and verb subcategorization probabilities. In P. Merlo & C. Stevenson (eds.), *Volume from CUNY-98*. Amsterdam: Benjamins.
Rosa, Eva, José L. Tapia & Manuel Perea. 2017. Contextual diversity facilitates learning new words in the classroom. *PloS one* 12(6). e0179004.
RStudio Team. 2015. *RStudio: Integrated Development Environment for R*. Boston, MA.
Rutvik, H. Desaia, Wonil Choib & John M. Henderson. 2018. Word frequency effects in naturalistic reading. *LANGUAGE, COGNITION AND NEUROSCIENCE* (2327-3798). 1–12.
Saffran, Jenny R., Richard N. Aslin & Elissa L. Newport (eds.). 1996. "*Statistical cues in language acquisition: Word segmentation by infants.*".
Schäfer, Michael. 2013. Phonetic Reduction of Adverbs in Icelandic: On the Role of Frequency and Other Factors. https://freidok.uni-freiburg.de/fedora/objects/freidok:9803/datastreams/FILE1/content. (17 July, 2018.)
Schmid, Hans Jörg. 2000. *English abstract nouns as conceptual shells.: From corpus to cognition*. Berlin - New York: Mouton de Gruyter.
Schmid, Hans Jörg. 2010. Entrenchment, Salience, and Basic Levels. In Geeraerts Dirk & Cuyckens Hubert (eds.), *The Oxford Handbook of Cognitive Linguistics*, 117–139. Oxford: Oxford Univ. Press.
Schmid, Hans Jörg. 2015. A blueprint of the entrenchment-and-conventionalization model. *Yearbook of the German Cognitive Linguistics Assosiation* 3(1). 3–26.
Schmid, Hans Jörg. 2017a. A Framework for Understanding Linguistic Entrenchment and Its Psychological Foundations. In Hans Jörg Schmid (ed.), *Entrenchment and the psychology of language learning.: How we reorganize and adapt linguisitc knowledge* (Language and the human lifespan), 9–35. Berlin: Walter de Gruyter.

Schmid, Hans Jörg (ed.). 2017b. *Entrenchment and the psychology of language learning.: How we reorganize and adapt linguistic knowledge* (Language and the human lifespan). Berlin: Walter de Gruyter.

Schmid, Hans Jörg. 2017c. Linguistic Entrenchment and Its Psychological Foundations. In Hans Jörg Schmid (ed.), *Entrenchment and the psychology of language learning.: How we reorganize and adapt linguistic knowledge* (Language and the human lifespan), 435–452. Berlin: Walter de Gruyter.

Schmid, Hans-Jörg. 2014. *Entrenchment, memory and automaticity. The psychology of linguistic knowledge and language learning*.

Schmid, Hans-Jörg. 2020. *The Dynamics of the Linguistic System: Usage, conventionalization, and Entrenchment.* Oxford: Oxford University Press.

Schmidt, Richard. 2010. Attention, awareness, and individual differences in language learning. 721–737.

Schmitt, Norbert (ed.). 2004. *Formulaic sequences.* Amsterdam: John Benjamins.

Schneider, Walter & Richard M. Shiffrin. 1977a. Controlled and Automatic Human Information Processing: I. Detection, Search and Attention. *Psychological Review* 84, https://cogs.sitehost.iu.edu/FestschriftForRichShiffrin/pubs/1977%20Controlled%20and%20Automatic%20Info%20Processing%20I.%20Shneider,%20Shiffrin.pdf. (8 March, 2021.)

Schneider, Walter & Richard M. Shiffrin. 1977b. Controlled and Automatic Human Information Processing: II. Perceptual Learning, Automatic Attending, and a General Theory. *Psychological Review* 84, https://cogs.sitehost.iu.edu/FestschriftForRichShiffrin/pubs/1977%20Controlled%20and%20Automatic%20Info%20Processing%20I.%20Shneider,%20Shiffrin.pdf. (8 March, 2021.)

Schuppler, Barbara, Mirjam Ernestus, Odette Scharenborg & Lou Boves. 2011. Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics* 39(1). 96–109.

Segalowitz, Norman. 2010. *Cognitive basis of second language fluency.* New York / London: Routledge.

Seidenberg, Mark S. & Maryellen C. MacDonald. 1999. A probabilistic constraints approach to language acquisition and processing. *Cognitive Science* 23(4). 569–5888.

Sinclair, John & Anna Mauranen. 2006. *Linear unit grammar: Integrating speech and writing / John McH. Sinclair, Anna Mauranen* (Studies in corpus linguistics, 1388-0373v. 25). Amsterdam, Philadelphia: J. Benjamins Pub. Co.

Siyanova-Chanturia, Anna. 2013. Neural Correlates of Lexical Processing. *The Mental Lexicon* 8, https://www.jbe-platform.com/content/journals/10.1075/ml.8.2.06siy.

Siyanova-Chanturia, Anna. 2015. On the 'holistic' nature of formulaic language. *Corpus Linguistics and Linguistic Theory* 11(2).

Skoog, Waller, Sara Mårten Eriksson & Patrik Sörqvist. 2015. Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. *Frontiers in psychology* 6, https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00978/full.

Smith, Nathaniel J. & Roger Levy. 2013. The effect of word predictability on reading time is logarithmic. *Cognition* 128(3). 302–319.

Song, Jieun. 2013. Variable schwa deletion in English: A corpus study. *Studies in Phonetics, Phonology and Morphology* 19(1). 35–52.

Squire, Larry R., Lisa Genzel, John T. Wixted & Richard G. Morris. 2015. Memory Consolidation. *Cold Spring Harbor perspectives in biology* 7(8). 1–21.
Staub, Adrian. 2015. The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistic Compass* 9(8). 311–327.
Stefanowitsch, Anatol & Susanne Flach. 2017. The Corpus-Based Perspective on Entrenchment. In Hans Jörg Schmid (ed.), *Entrenchment and the psychology of language learning.: How we reorganize and adapt linguisitc knowledge* (Language and the human lifespan), 101–129. Berlin: Walter de Gruyter.
Stoet, Gijsbert. 2010. PsyToolkit: a software package for programming psychological experiments using Linux. *Behavior Research Methods* 42, https://link.springer.com/article/10.3758/BRM.42.4.1096.
Stoet, Gijsbert. 2017. PsyToolkit. *Teaching of Psychology* 44(1). 24–31.
Stone, James M. & John N. Towse. 2015. A Working Memory Test Battery: Java-Based Collection of Seven Working Memory Tasks. *Journal of Open Research Software* 3, http://dx.doi.org/10.5334/jors.br. (2 June, 2020.)
Street, James & Ewa Dąbrowska. 2014. Lexically specific knowledge and individual differences in adult native speakers' processing of the English passive. *Applied Psycholinguistics* 35. 97–118.
Street, James A. & Ewa Dąbrowska. 2010. More individual differences in language attainment: How much do adult native speakers of English know about passives and quantifiers? *Lingua* 120(8). 2080–2094.
Tabossi, Patrizia, Rachele Fanari & Kinou Wolf. 2009. Why are idioms recognized fast? *Memory & Cognition* 37(4). 529–540.
Takashima, Atsuko & Iske Bakker. 2017. Memory Consolidation. In Hans Jörg Schmid (ed.), *Entrenchment and the psychology of language learning.: How we reorganize and adapt linguistic knowledge* (Language and the human lifespan), 177–201. Berlin: Walter de Gruyter.
Tanaka-Ishii, Kumiko & Hiroshi Terada. 2011. Word familiarity and frequency. *Studia Linguistica* 65(1). 96–116.
Taylor, John R. 2012. *The mental corpus: How language is represented in the mind.* Oxford: Oxford University Press.
Theakston, Anna L. 2017. Entrenchment in First Language Learning. In Hans Jörg Schmid (ed.), *Entrenchment and the psychology of language learning.: How we reorganize and adapt linguisitc knowledge* (Language and the human lifespan), 315–343. Berlin: Walter de Gruyter.
Tomasello, Michael. 2005. *Constructing a language: A usage-based theory of language acquisition / Michael Tomasello.* Cambridge, Mass., London: Harvard University Press.
Tremblay, Antoine, Bruce Derwing, Gary Libben & Chris Westbury. 2011. Processing Advantages of Lexical Bundles: Evidence From Self-Paced Reading and Sentence Recall Tasks. *Language Learning* 61(2). 569–613.
Trouvain, Jürgen. 2004. *Tempo Variation in Speech Production. Implications for Speech Sythesis.* (PHONUS 8). Saarbrücken, Germany: Institute of Phonetics, Saarland University.
Trueswell, John C. & Michael K. Tanenhaus. 1994. Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In Jr. Clifton & C. Frazier (eds.), *Perspectives on Sentence Processing*, 155–179. Hillsdale, NJ: Lawrence Erlbaum.

Tsao, Y. C. & G. Weismer. 1997. Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component. *Journal of Speech, Language, & Hearing Research* 40(4). 858–866.

Ullman, Michael T. 2004. Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92, https://www.sciencedirect.com/science/article/pii/S0010027703002324.

Ullman, Michael. T., Robin. A. Miranda & Michelle. L. Travers. 2007. Sex Differences in the Neurocognition of Language. In Jill B. Becker, Karen J. Berkley, Nori Geary, Elizabeth Hampson, James P. Herman & Elizabeth Young (eds.), *Sex Differences in the Brain*, 291–311. Oxford: Oxford University Press.

van Berkum, Jos J. A., Colin M. Brown, Pienie Zwitserlood, Valesca Kooijman & Peter Hagoort. 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of experimental psychology. Learning, memory, and cognition* 31(3). 443–467.

van Dyke, Julie A. & Clinton L. Johns. 2012. Memory Interference as a Determinant of Language Comprehension. *Language and Linguistics Compass* 6(4). 193–211.

van Dyke, Julie A. & Brian McElree. 2006. Retrieval interference in sentence comprehension. *Journal of Memory and Language* 55, https://www.sciencedirect.com/science/article/pii/s0749596x0600043x?casa_token=bnfbmv2hrq0aaaaa:vulianscvgrwq1ha1zcx7c7fyxqfeipntpmprfc0vjmnuvganccfwtojzqtr1brfgna-jer0fq.

van Petten, Cyma & Barbara J. Luka. 2012. Prediction during language comprehension: benefits, costs, and ERP components. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology* 83(2). 176–190.

Vandierendonck, André, Baptist Liefooghe & Frederick Verbruggen. 2010. Task switching: interplay of reconfiguration and interference control. *Psychological bulletin* 136(4). 601–626.

Vergara-Martínez, Marta, Montserrat Comesaña & Manuel Perea. 2017. The ERP signature of the contextual diversity effect in visual word recognition. *Cognitive, Affective, & Behavioral Neuroscience* 17, https://link.springer.com/article/10.3758/s13415-016-0491-7.

Verhagen, Véronique. 2020. *Illuminating variation: Individual differences in entrenchment of multi-word units*. Netherlands: Tilburg University Doctoral Thesis.

Verkoeijen, Peter P. J. L., Remy M. J. P. Rikers & Henk G. Schmidt. 2004. Detrimental influence of contextual change on spacing effects in free recall. *Journal of experimental psychology. Learning, memory, and cognition* 30(4). 796–800.

Warker, Jill A., Gary S. Dell, Christine A. Whalen & Samantha Gereg. 2008. Limits on learning phonotactic constraints from recent production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34(5). 1289–1295.

Warner, Natasha & Benjamin V. Tucker. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America* 130(3). 1606–1617.

Warren, Paul. 2013. *Introducing psycholinguistics.* Cambridge: Cambridge University Press.

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

Wilson, Michael. 1988. MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods* 20(1). 6–10.
Yaneva, Victoria, Shiva Taslimipoor, Omid Rohanian & an Le Ha. 2017. Cognitive Processing of Multiword Expressions in Native and Non-native Speakers of English: Evidence from Gaze Data. 363–379.
Ye, Zheng & Xiaolin Zhou. 2009. Executive control in language processing. *Neuroscience and biobehavioral reviews* 33, https://www.sciencedirect.com/science/article/pii/S0149763409000384.
Yoshimura, Yuki & Brian MacWhinney. 2007. The effect of oral repetition on L2 speech fluency: an experimental tool and language tutor. *undefined*, https://www.semanticscholar.org/paper/The-effect-of-oral-repetition-on-L2-speech-fluency%3A-Yoshimura-MacWhinney/78629c034358c2893fe58580e31c24ea43ed5a68.
Yuan, Jiahong,, Jiahong, Mark Liberman & Christopher Cieri. 2006. Towards an integrated understanding of speaking rate in conversation. *Ninth International Conference on Spoken Language Processing*.
Zarcone, Alessandra, Marten van Schijndel, Jorrig Vogels & Vera Demberg. 2016. Salience and Attention in Surprisal-Based Accounts of Language Processing. *Frontiers in Psychology* 7. 1–17.
Zellers, Margaret, Barbara Schuppler & Meghan Clayards. 2018. Introduction, or: why rethink reduction. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler & M. Zellers (eds.), *Rethinking Reduction: Interdisciplinary Perspectives on Conditions, Mechanisms, and Domains for Phonetic Variation* (Phonology and Phonetics), 1–24. Berlin/Boston: DE GRUYTER MOUTON.
Zimmerer, Frank & Henning Reetz. 2011. Same or Different? Schwas in Naturally Spoken German. *City University of Hong Kong. Department of Chinese, Translation & Linguistics*. 2324–2327.

12 Appendix

Appendix 1. Stimulus text used for experiment I. The target items are boldly typed

Urteil im Verfahren wegen unerlaubter Verwendung von Pestiziden ergangen

Die Geschichte begann im Juni 2018 in Bohl, einer Kleinstadt in der Westukraine, als ein 44 - jähriger Bewohner eines nahegelegenen Dorfes im Krankenhaus starb. Der Patient war wegen unklarer Beschwerden ins Krankenhaus eingeliefert worden, sein Zustand verschlechterte sich jedoch rapide und er konnte nicht gerettet werden. Die Familie vermutete eine Chemikalienvergiftung, die auf den Einsatz bestimmter Pestizide bei der Feldarbeit zurückzuführen sei. Außerdem kam es in dieser Gegend zu weiteren gesundheitlichen Beschwerden bei Menschen und einer ungewöhnlich hohen Zahl von Todesfällen bei Haustieren. Als Folge dieser Ereignisse hat die zuständige **Staatsanwaltschaft** Lemberg ein **Ermittlungsverfahren** gegen den Hersteller der Pestizide eingeleitet. Rund zwei Monate lang hat die **Staatsanwaltschaft** umfangreiche **Ermittlungsmaßnahmen** in der Sache durchgeführt, um dem Verdacht nachzugehen, dass das Großunternehmen für den Todesfall verantwortlich sei. Aufgrund zahlreicher Bestechungsversuche seitens des Konzerns mit dem Ziel den Fall beizulegen, haben die betroffenen Familien ein **Klageerzwingungsverfahren** beantragt. Danach wurden mehr als 60 Bewohner aus Nachbardörfern befragt, um den Verdacht zu erhärten. Nach dieser Feststellung erhebt die **Staatsanwaltschaft** den Vorwurf gegen den Hersteller, dass sich der Konzern der mit der Verwendung des Pestizids verbundenen Risiken bewusst gewesen sein muss. Der Hersteller berief sich auf sein **Zeugnisverweigerungsrecht**, weil sich der Konzern mit Aussagen während des Verfahrens selbst belasten könnte. Dieser Schritt des Unternehmens hat das Interesse des **Strafverfolgungsbehörden** geweckt. "Wir sind der Ansicht, dass es kein absolutes **Verfahrenshindernis** darstellt, das zu einer **Verfahrensverzögerung** führen könnte", sagte der Sprecher der Lemberger **Staatsanwaltschaft**, Ivan Pavlov. Ende 2010 hat der oberste Verfassungsgerichtshof der Ukraine, der dem **Bundesverfassungsgericht** entspricht, in einem bemerkenswerten Urteil festgestellt, dass die Verwendung von Chemikalien, die ein Gesundheitsrisiko für Menschen darstellen könnten, im Umkreis von 100 km von bewohnten Gebieten verboten ist. Die betroffenen Felder des Konzerns sind 5 km von Bohl entfernt. Am 12. September fand die **Beschuldigtenvernehmung** statt: 10 Mitarbeiter wurden verhaftet, einschließlich des Unternehmensleiters sowie Mitarbeitern der Behörde, die die Genehmigung zu Verwendung der schädlichen Chemikalien erteilt hatte. Der Konzern muss die Kosten des **Rechtsmittelverfahrens** tragen und allen Betroffenen eine Entschädigung von umgerechnet 1500 Euro bezahlen. Als Begründung wurde angeführt, dass das Unternehmen gegen § 78 Abs. 52 **Gerichtsverfassungsgesetz** verstoßen hat. Gleichzeitig erließ das Gericht ein **Beweisverwertungsverbot**. Der Fall wurde damit abgeschlossen

Appendix 2. Stimuli used for experiment II

Item	Type	FreqDWDS	FreqLect	FreqWeb	FreqBase	FreqStEnding	Dispersion	TPComp
Arbeitsmittel	general	118		28591	7491377	11275406	65	0,0025357
Auseinandersetzung	general	4829		693022	693022	729081	2010	0,95054185
Bekantmachung	general	808		67446	487913	961485	459	0,07014774
Beschäftigungsmöglichkeit	general	39		16950	413671	258914	38	0,06546575
Bevölkerungsgruppe	general	224		56546	1376466	286498	161	0,19736962
Diskriminierung	general	785		180662	180662	210281	10528	0,85914562
Eigenständigkeit	general	271		56622	11429973	3586941	4902	0,0157856
Eigenverantwortung	general	67		73793	11429973	3586941	1096	0,02057268
Einwanderungsbehörde	general	13		3597	484192	46487	322	0,07737647
Eisenbahnverwaltung	general	325		1280	194309	252041	67	0,00507854
Fußballweltmeisterschaft	general	18		16917	770452	1206422	1062	0,01402246
Gemeinschaftsarbeit	general	181		9181	801340	492900	306	0,0186265
Gemeinschaftsproduktion	general	32		3841	801340	492900	194	0,00779266
Gesellschafterversammlung	general	26		16721	2918330	610449	277	0,02739131
Hochschulstudium	general	105		21830	989581	1515478	379	0,0144047
Industriegesellschaft	general	378		16057	796879	920740	2012	0,01743923
Informationsspeicherung	general	7		694	5544890	1648986	12	0,00042087
Lächerlichkeit	general	302		15290	222500	26694	978	0,57278789
Lebensmittelgeschäfte	general	73		16173	27777	449045	441	0,03601644
Lichtgeschwindigkeit	general	103		25143	1920801	1475609	370	0,01703907
Liebesgeschichte	general	191		81108	3148687	855699	2119	0,09478567
Marktwirtschaft	general	734		118429	2862321	2055455	5	0,05761693
Meinungsfreiheit	general	205		76983	2784440	410653	3194	0,18746484
Musikwissenschaftler	general	64		8021	3144142	2974202	425	0,00269686
Privatunterricht	general	99		7560	2547806	1905944	115	0,00396654
Schulterblätter	general	203		13864	490249	203609	270	0,06809129
Sozialwissenschaftler	general	84		15286	3378694	2695713	1315	0,00567049
Temperaturunterschiede	general	53		15571	942209	234025	188	0,06653563

Verdienstmöglichkeit	general	54		16584	236544	78408	358	0,21150903
Wirtschaftskorrespondentÿ	general	21		283	1571805	2626745	38	0,00010774
Anklageerhebung	terminology	44	33	4787	118759	49172	703	0,09735215
Beschlagnahme	terminology	654	97	16793	48431	40246	350	0,41725886
Beschuldigtenvernehmung	terminology	1	3	1034	77084	1978	6	0,52275025
Beweisverwertungsverbot	terminology	1	263	3401	685418	263305	12	0,01291658
Beweiswürdigung	terminology	53	74	12260	685418	263305	12	0,04656197
Bundesverfassungsgericht	terminology	866	48	187348	187348	198105	1	0,94570051
Ermittlungsmaßnahmen	terminology	5	35	3346	427266	146697	36	0,02280892
Ermittlungsrichter	terminology	42	62	5302	427266	146697	623	0,03614253
Ermittlungsverfahren	terminology	203	260	49840	427266	146697	4055	0,33974792
Gerichtsverfassungsgesetz	terminology	100	10	2503	1416515	2723	82	0,91920676
Grundrechtseingriffe	terminology	2	22	4323	155146	34106	59	0,12675189
Hauptverfahren	terminology	85	88	7149	233477	5431687	520	0,00131617
Hauptverhandlung	terminology	501	607	34132	233477	5431687	1087	0,00628387
Klageerzwingungsverfahren	terminology	8	4	727	461800	270780	13	0,00268484
Rechtsmittelverfahren	terminology	10	6	3445	3312396	4018487	18	0,00085729
Revisionsbegründung	terminology	21	7	2379	198596	103614	13	0,02296022
Revisionsgericht	terminology	66	27	6235	198596	103614	57	0,06017527
Staatsanwaltschaft	terminology	1540	505	291051	2498618	2577162	25594	0,11293469
Strafprozessordnung	terminology	282	14	12773	392120	1779460	643	0,00717802
Strafvereitelung	terminology	16	2	3347	392120	1779460	294	0,00188091
Strafverfahrensrecht	terminology	10	31	1067	392120	1779460	20	0,00059962
Strafverfolgungsbehörden	terminology	46	80	19569	392120	1779460	744	0,01099716
Verfahrensbeteiligten	terminology	12	43	8322	1575320	257511	138	0,03231707
Verfahrenshindernis	terminology	2	37	1397	1575320	257511	29	0,00542501
Verfahrensrüge	terminology	3	17	5270	1575320	257511	4	0,02046515
Verfahrensverzögerung	terminology	1	7	2572	1575320	257511	16	0,00998792
Wahrheitsfindung	terminology	65	19	7826	1015255	77021	523	0,10160865
Zeugnisverweigerungsrecht	terminology	52	85	4837	321186	38181	143	0,12668605

Zwangsmaßnahmen	terminology	171	138	13897	207076	423783	678	0,03279273
Zwischenverfahren	terminology	1	31	941	1575320	1049432	13	0,00089668

Appendix 3.1. Stimuli sentences for experiment III. The target collocations are in bold

Sentence	Word1	Word2	Word3	Word4	Word5	Probe
Wann das Landgericht den Termin zur **mündlichen Verhandlung** ansetzt, ist bislang nicht bekannt.	schriftliche	ausdrückliche	anstehende	sachliche	schwierige	schriftliche
Die Temperaturen auf der **ganzen Welt** haben letztes Jahr neue Rekordwerte erreicht.	vergangenes	kommendes	folgendes	voriges	schwieriges	voriges
Der Angeklagte hat die gegen ihn **verhängte Strafe** widerspruchslos angenommen.	gerechte	saftige	milde	strenge	verdiente	stärke
Sie verfügt über ein **fotografisches Gedächtnis** und kann logisch denken.	kollektives	kulturelles	visuelles	prozedurales	implizites	gesellschaftliches
Es ist eine echte Herausforderung für jeden **erfahrenen Ermittler**, einen Serienmörder zu fassen.	verdeckter	sympatischer	forensischer	schrulliger	leitender	festgelegter
Der Nachweis bescheinigt, wie und wo Strom aus **erneuerbaren Energien** produziert wurde.	regenerative	elektrische	kinetische	erzeugte	saubere	positive
Der Angeklagte wurde wegen unerlaubten Besitzes von Betäubungsmitteln in nicht **geringer Menge** verurteilt.	ausreichende	niedrige	erhebliche	deutliche	kleine	große

In diesem Jahr fand der Kurs zum **ersten Mal** auf der Marienburg statt.	letztes	einziges	unzähliges	allererstes	dutzendes	zweites
Das Gericht verurteilte den Mann wegen **versuchten Mordes** zu einer Freiheitsstrafe von dreieinhalb Jahren.	brutaler	kaltbültiger	begangener	verübter	ungeklärter	verübter
Asthma ist heutzutage keine **schreckliche Krankheit** mehr, man kann viel dagegen tun.	unheilbare	schlimme	traurige	schwere	tragische	traurige
Der Gutachter konnte keine Zeichen für **eingeschränkte Schuldfähigkeit** erkennen.	verminderte	reduzierte	geringe	erhöhte	übermaßige	niedrige
Die Module dieses Studiengangs können in **beliebiger Reihenfolge** angeordnet werden.	alphabetische	gewünschte	gleiche	verschiedene	entsprechende	unsortierte
Im Augenblick sind im **vorliegenden Fall** die Voraussetzungen für eine Verurteilung nicht gegeben.	entscheidender	schwerer	konkreter	deutlicher	vorgelegter	entscheidender
Das klingt logisch, aber auf der **andere Seite** auch sehr kompliziert.	rechte	folgende	dunkle	schöne	sichere	dunkle
Er steht unter **dringendem Tatverdacht**, die Straftat begangen zu haben.	hinreichender	rechtfertigender	begründeter	fortbestehender	genügender	kluger
Die Zahl der jährlich von Jugendlichen **verübten Straftaten wurde** mit über 20 000 beziffert.	begangene	geschehene	unaufgeklärte	schändliche	gemachte	unaufgeklärte
Ich halte diesen **ohrenbetäubenden Lärm** nicht mehr aus.	störender	unerträglicher	tosender	verursachter	nächtlicher	stiller

Man sollte viel trinken und auf **gesunde Ernährung** achten mit viel Gemüse und Obst.	ausgewogene	gute	bewusste	einfache	frische	alte
Der Senat kann auf der Grundlage der bisher **getroffenen Feststellungen** diese Frage nicht entscheiden.	tatrichterliche	bindende	gerichtliche	gesonderte	angegriffene	besondere
Ich halte es für keine **gute Idee**, mit dem Auto zu fahren.	kreative	tolle	geniale	super	schlechte	schlimme
Das Ziel der Ermittler ist, **entlastende Beweise** zu finden.	handfeste	stichhaltige	eindeutige	schlagende	endültige	vor
Für Deutschland kann der **demografische Wandel** zu einem großen Problem werden.	stetiger	tiefgreifender	technologischer	gesellschaftlicher	grundlegender	technologischer
Befindet man sich in einer **ausweglosen Lage**, kann man seine Freunde um Hilfe bitten.	missliche	aussichtlose	unlösbare	trostlose	brenzlige	aussichtlose
Was ihr besonders gefiel, war das **sprudelnde Wasser** im Garten.	kaltes	warmes	sauberes	fließendes	kochendes	fließendes
Die vom Berufungsgericht **zugelassene Revision** der Beklagten hatte Erfolg.	eingelegte	gerichtete	eingeschränkte	gestützte	gründliche	eingelegte
Zum Glück gibt es ja inzwischen **andere Quellen** zu diesem Thema.	heiße	zuverlässige	seriöse	schriftliche	historische	historische

Dieses Urteil entspricht der **ständigen Rechtsprechung** des Bundesverfassungsgerichts, an der der Senat festhält.	höchstrichterliche	bundesgerichtlich	gefestigte	ergangene	bischerige	vorherige
Im Mülleimer hat er nur **zusammengeknülltes Papier** gefunden.	bedrucktes	gefälschtes	wertloses	dünnes	farbiges	dickes
Ein Vertrag erfordert laut **geltender Rechtsprechung** zwei gleichlautende Willenserklärungen.	ständige	gefestigte	bisherige	gültige	festgelegte	gefestigte
Nach vielmaligem Sonnenbaden sehen Sie nicht nur aus wie **zerknittertes Papier**, sondern riskieren Sie auch noch Hautkrebs.	zusammengeknülltes	zerrissenes	gebügeltes	farbiges	eingewickeltes	glattes
Gegen dieses Vorgehen muss **sofortige Beschwerde** eingelegt werden.	körperliche	staatsrechtliche	gesundheitliche	erhobene	typische	richterliche
Ich sehe nicht viel fern, aber **letzte Woche** waren die Tennisübertragungen.	vergangene	kommende	ganze	vorige	erste	folgende
Feindliche Spione halten den Mann für einen **verdeckten Ermittler** des Geheimdienstes.	polizeilicher	kauziger	leitender	sympathischer	eigenwilliger	polizeilicher
Mit ihren blonden, **schulterlangen Haaren** sah sie unheimlich hübsch aus.	graues	dunkles	lockiges	glattes	gefärbtes	glattes
Von geisteskranken Tätern **verübte Morde** machen den weitaus größten Teil aller Tötungsdelikte aus.	vorbestrafter	engagierter	mutiger	aktiver	offensichtlicher	passiv

Bis 2020 sollte der Anteil an **sauberer Energie** sogar auf rund 45 Prozent steigen.	richtige	einfache	erneuerbare	komplette	schnelle	schnelle
Das Gericht verurteile den Angeklagten wegen **gefährlicher Körperverletzung** zu drei Jahren Haft.	fahrlässige	strafbare	begangene	schwere	absichtliche	schwere
Die Sitzung dauerte den **ganzen Tag,** und trotzdem sind noch Fragen ungeklärt.	nächster	schöner	erster	vergangener	selber	schöner
Der Rechtsanwalt plädierte auf **verminderte Schuldfähigkeit,** und somit auf Freispruch.	ausschießbare	aufgehobene	erhebliche	strafrechtliche	fehlende	gültige
Zugegeben, bei den **drakonischen Strafen** ist man weit über das Ziel hinausgeschossen.	saftige	milde	verhängte	strenge	verdiente	verhängte
Sie bemerkte, dass er ein **phänomenales Gedächtnis** für Fakten hat.	fotographisches	sensationelles	fabelhaftes	grandioses	beeindruckendes	typisches
Im Schwarzwald gibt es viele Quellen mit schönem **klarem Wasser**, das man trinken kann.	fließendes	sprühendes	prickelndes	sprudelndes	erfrischendes	stilles
Der Landesgerichtshof hat die **weitergehende Revision** der Kläger zurückgewiesen.	prozessuale	gründliche	anhängige	beschränkte	erfolglose	saubere
Nur wenige wissen, dass dieses Gebiet vor **langer Zeit** unbewohnt war.	kurze	heutige	gleiche	damalige	frühe	gestrige
Der Anteil der von Jugendlichen **begangenen**	motivierte	rechtsextreme	vorsätzliche	terroristische	vorgeworfene	vorsätzliche

Straftaten liegt knapp unter 50 Prozent.						
Überall auf der Welt gibt es Menschen, die an **unheilbaren Krankheiten** leiden.	chronische	schwere	tödliche	seltene	schlimme	schwere
Viele Menschen achten nicht genügend auf eine **ausgewogene Ernährung** und essen viel zu viel Schokolade.	vegane	ungesunde	vollwertige	falsche	bewusste	pflanzliche
Störungen der Nachbarn durch **unnötigen Lärm**, laute Musik etc. sind zu vermeiden.	ohrenbetäubender	lauter	tosender	donnernder	lautstarker	ohrenbetäubender
Das Berufungsurteil hält der **rechtlichen Nachprüfung** in einem entscheidenden Punkt nicht stand.	mündliche	bestandene	gebotene	amtliche	beschränkte	beschränkte
Sie sollen sich eine **eigene Meinung** bilden und sie vertreten können.	einhellige	öffentliche	abweichende	persönliche	ehrliche	abweichende
Das Motiv des bereits **vorbestraften Täters** bleibt unklar.	unbekannter	mutmaßlicher	maskierter	jugendlicher	potentieller	männlicher
Der Staatsanwalt muss bei **begründetem Tatverdacht** einer strafbaren Handlung ermitteln.	dringender	hinreichender	beruhender	konkreter	solcher	konkreter
Er öffnete mir die **verriegelte Türe** zur Zelle, die in tiefer Dunkelheit lag.	verschlossene	versperrte	offenstehende	einbruchsichere	klemmende	geöffnete

Vor Gericht hat jeder Mann Anspruch auf **rechtliches Gehör** und auf ein faires Verfahren.	geschultes	lautstarkes	empfindliches	intaktes	absolutes	lautes
Das ist natürlich in **erster Linie** ihr eigenes Verdienst.	gerade	klare	ganze	feine	rote	blaue
Er hat damit im Ergebnis das **angefochtene Urteil** des Finanzgerichts bestätigt.	rechtskräftiges	erstinstanzliches	ergangenes	vernichtendes	veröffentlichtes	erstinstanzliches
Diese Rechtssache hat keine **grundsätzliche Bedeutung**, also müssen wir sie nicht weiter verfolgen.	besondere	entscheidende	zunehmende	große	wachsende	besondere
Besonders auf dem Land gibt es ein **großes Problem** mit öffentlichem Nahverkehr.	technisches	eigentliches	gleiches	erhebliches	änliches	erhebliches
Der Richter hat die **erstinstanzliche Entscheidung** bestätigt.	endgültige	getroffene	richtige	politische	strategische	schlimme
Plötzlich standen sie vor zwei **verschlossenen Türen** und konnten nur noch zurückgehen.	geöffnete	breite	schmale	knarrende	zweiflügelige	zweiflügelige
Am Samstagabend brachen zwei **männliche Täter** in ein Wohnhaus ein.	vorbestrafter	verurteilter	geständiger	entlassener	verhafteter	vorheriger
Angeblich gebe es dafür **stichhaltige Beweise**, die man der Öffentlichkeit aber bislang vorenthalten hat.	entlastende	eindeutige	derartige	klare	berechtigte	mutige
Im Hochwasserfall werden die **zuständigen Behörden** automatisch alarmiert.	staatliche	belangte	örtliche	lokale	beteiligte	ruhige

So ist es im **echten Leben**, ob es dir nun passt oder nicht.	tägliches	ewiges	menschliches	normales	ganzes	ewiges
Am häufigsten stützt sich der **erlassene Haftbefehl** auf eine mögliche Fluchtgefahr.	antragsgemäßer	erwirkter	vollstreckter	ergangener	richterlicher	erwirkter
Ich kann mich aus dieser **misslichen Lage** nur selbst befreien.	ruhige	zentrale	geographische	finanzielle	schwierige	leichte
Nach einem heute **ergangenen Urteil** muss Google persönlichkeitsverletzende Suchvorschläge entfernen.	angefochtenes	ergangenes	angegriffenes	erlassenes	streitiges	angegriffenes
Die Kollegien in den **naturwissenschaftlichen Schulen** zählen zu den ältesten in Europa.	vertiefende	weiterführende	ergänzende	praxisnahe	fachliche	normale
Die Anwendbarkeit neuer Prozessgesetze auf **anhängige Rechtsstreite** ist umstritten.	langwierige	vorliegend	jahrelange	andauernde	erbitterte	vorliegende
Ein erheblicher Einfluss wird auch den **sozialen Netzwerken** zugeschrieben.	lokales	drahtloses	globales	neuronales	virtuelles	technisches
Das Strafgesetzbuch sieht für Mord eine **lebenslange Freiheitsstrafe** vor.	verhängte	zeitige	ausgesetzte	rechtskräftige	mehrjährige	zeitige
Nach dem Wechsel auf eine **weiterführende Schule** sind manche Kinder viel stärker motiviert.	berufsbildende	staatliche	öffentliche	allgemeine	praxisnahe	berufsbildende
Erst Ende November soll die **endgültige Entscheidung** bekannt gegeben werden.	erstinstanzliche	definitive	offenbare	schnelle	falsche	richtige

Vor der Abfahrt des Schiffes kam es zu einem **schmerzlichen Abschied** von allen Verwandten und Freunden.	tränenreicher	herzzerreißender	rührseliger	schmerzvoller	emotionsgeladener	rührseliger
Außerdem braucht Entbürokratisierung neben dem **rechtlichen Rahmen** vor allem eine Vision.	zeitlicher	festlicher	angemessener	passender	institutioneller	finanzieller
An dieser Stelle noch einmal **herzlichen Glückwunsch** zu dieser Leistung.	nachträglicher	allerbester	aufrichtiger	verspäteter	persönlicher	nachträglicher
Nach den Worten der ermittelnden Staatsanwaltschaft sei ein **fremdenfeindlicher Hintergrund** der Tat nicht auszuschließen.	zuständige	anklagende	überlastete	beschwerdeführend	lokale	ständige
Soziologen haben festgestellt, dass der **gesellschaftliche Wandel** immer schneller erfolgt.	demographischer	sozialer	kultureller	ökonomischer	wirtschaftlicher	ökonomischer
Ein von einem deutschen Gericht **ergangener Haftbefehl** kann allerdings weltweit vollstreckt werden.	erlassener	völliger	erneuter	rascher	definitiver	rascher
Der Mandant ist zu einer **mehrjährigen Freiheitsstrafe** verurteilt worden.	lebenslange	jahrelange	langfristige	dauernde	zwangsläufige	kurze
Am Bahnhof nahm die Mutter **tränenreichen Abschied** von ihrem Sohn, was ihm sehr peinlich war.	wehmütiger	schmerzlicher	trauriger	bewegender	rührender	froher
Es fiel ihm sehr schwer die Wörter in die **alphabetische Reihenfolge** zu bringen.	chronologische	aufsteigende	lose	festgelegte	zufällige	aufsteigende

Mit seinen **schmutzigen Haaren** sah er nicht sehr nett aus.	schulterlanges	dreckiges	sauberes	verschmutztes	dunkles	verschmutztes
Die Anordnung war von der **zuständigen Staatsanwaltschaft** beantragt worden.	ermittlende	verantvortliche	tätige	beteiligte	betreffende	ermittlende

Appendix 3.2. Stimuli used for experiment III and the linguistic predictors

Item	**ItemType**	**Condition**	**Adjective**	**Freq_Web**	**MI_Web**	**T.Score_Web**	**X2_Web**	**DP_Web**	**LL_Web**	**TP_Web**
Abschied	general	high	tränenreicher	294452	7,97052685	19,0027194	90580,3137	0,01080147	3304,16958	1,0844306
Abschied	general	low	schmerzlichen	294452	12,8393648	33,0106432	7986606,78	0,31681842	17640,4887	31,686046
Energie	general	high	erneuerbare	2280519	9,62809434	610,466464	295355631	0,85522437	4905396,73	85,589956
Energie	general	low	saubere	2280519	4,62077093	81,3021349	162851,654	0,02553731	32451,4871	2,6611433
Ernährung	general	high	ausgewogene	754383	8,24618544	330,493525	33210793,1	0,12014953	1088377,35	12,042672
Ernährung	general	low	gesunde	754383	9,18937821	196,675235	22595886,5	0,23119953	430985,606	23,155395
Gedächtnis	general	high	fotografisches	236027	8,78832717	39,8596108	702539,935	0,02462467	16349,141	2,4678758
Gedächtnis	general	low	phänomenales	236027	8,38717555	19,8903505	132467,886	0,01863269	3844,48056	1,8688078
Glückwunsch	general	frequency	herzlich	345435	10,1966402	425,949991	213128158	0,25756486	2528377,33	25,758651
Haar	general	high	schulterlanges	1134978	10,8357097	50,9622946	4748675,83	0,83879474	37251,7625	83,925113
Haar	general	low	schmutzig	1134978	0,97831931	3,84599579	29,1576439	0,00044555	22,6697073	0,0904735
Idee	general	frequency	gute	3631331	3,32764958	423,812007	1838095,78	0,0125703	659778,482	1,3715340
Jahr	general	frequency	letztes	36081680	4,55019589	1260,08526	37878161,4	0,19113239	8340902,15	19,774985
Krankheit	general	high	unheilbare	1254709	4,6250431	38,8912976	37344,7342	0,00898625	7402,89669	0,9364011
Krankheit	general	low	schrekliche	1254709	10,0130491	83,3578509	7182810,04	0,39174299	86042,8714	39,211489
Lage	general	low	missliche	2578664	7,00599051	35,2063325	159451,453	0,10806406	9871,72543	10,891003
Lage	general	low	ausweglose	2578664	9,6288672	112,97653	10114198,7	0,67002744	157075,357	67,086085
Lärm	general	high	ohrenbetäubender	217206	12,7891746	58,1295538	23918933,9	0,27482863	54525,6718	27,486378
Lärm	general	low	unnötiger	217206	6,28093757	25,9863485	52525,0548	0,00298147	4681,18321	0,3019555

Leben	general	frequency	echtes	8671080	2,51912263	121,451429	84901,985	0,01109844	40196,5061	1,3420076
Linie	general	frequency	erste	1553632	5,16495057	770,594513	21697081,3	0,03973623	3781844,65	4,0179654
Mal	general	frequency	erstes	4145949	4,6162018	1002,30749	25136975,3	0,06813003	5466222,66	6,9816799
Meinung	general	frequency	eigene	2784440	3,50846801	234,163235	631298,284	0,00602833	207893,508	0,6536528
Netzwerk	general	frequency	sozial	1275074	7,19464428	488,481877	35077351,5	0,09086256	2074240,25	9,1222976
Papier	general	high	zusammengeknülltes	873123	9,46816169	10,801382	82649,8567	0,13413883	1318,53813	13,432835
Papier	general	high	zusammengeknülltes	873123	7,97071281	12,5592541	39576,3097	0,04738682	1447,94996	4,7576301
Problem	general	frequency	großes	6869925	2,60088263	406,042692	1023723,9	0,01045337	481800,883	1,2253712
Quelle	general	frequency	andere	1571190	2,1414915	141,270114	89712,2374	0,00158727	49143,8152	0,2015457
Reihenfolge	general	high	alphabetische	266474	6,72867074	59,3681818	373958,776	0,0127808	26549,8826	1,2898546
Reihenfolge	general	low	beliebiger	266474	11,7709742	106,430797	39592813,1	0,42492634	169171,802	42,503562
Schule	general	high	weiterführende	4321027	-1,3451901	-9,74384246	37,4082778	-0,0005545	48,6964778	0,0359841
Schule	general	low	naturwissenschaftliche	4321027	8,10212235	229,42057	14479167,4	0,2503527	507755,17	25,12094
Seite	general	frequency	andere	8954459	3,6224444	738,676561	6865170,91	0,03652714	2212993,01	3,9038523
Tag	general	frequency	ganzer	14738939	3,44772705	534,425063	3154512,44	0,04545458	1064633,7	4,9661068
Türe	general	high	verschlossene	1593032	9,3615144	22,7469387	340478,495	0,22079776	5823,89463	22,113336
Türe	general	low	verriegelte	1593032	10,2911811	179,269746	40285647,5	0,4209216	414359,74	42,122274
Wandel	general	high	demographischer	483459	10,6746545	197,661706	63884462	0,43973156	527704,913	43,995816
Wandel	general	low	gesellschaftlicher	483459	5,86844182	118,101503	815904,067	0,01547154	90129,1569	1,5725236
Wasser	general	high	sprudelndes	4004467	6,6028804	35,9580192	125790,576	0,08508248	9586,48144	8,5965483
Wasser	general	low	klares	4004467	4,59529234	156,336164	592123,914	0,020522	119773,494	2,1378630
Woche	general	frequency	letzte	7093123	4,9370403	709,261995	15589912,6	0,05992627	2836044,22	6,1357627
Zeit	general	frequency	lange	16945136	4,92586476	902,39393	25032922,8	0,18147739	4558540,22	18,669798
Abschied	general	high	tränenreicher	294452	7,97052685	19,0027194	90580,3137	0,01080147	3304,16958	1,0844306
Abschied	general	low	schmerzlichen	294452	12,8393648	33,0106432	7986606,78	0,31681842	17640,4887	31,686046
Energie	general	high	erneuerbare	2280519	9,62809434	610,466464	295355631	0,85522437	4905396,73	85,589956
Energie	general	low	saubere	2280519	4,62077093	81,3021349	162851,654	0,02553731	32451,4871	2,6611433
Ernährung	general	high	ausgewogene	754383	8,24618544	330,493525	33210793,1	0,12014953	1088377,35	12,042672
Ernährung	general	low	gesunde	754383	9,18937821	196,675235	22595886,5	0,23119953	430985,606	23,155395
Gedächtnis	general	high	fotografisches	236027	8,78832717	39,8596108	702539,935	0,02462467	16349,141	2,4678758

Gedächtnis	general	low	phänomenales	236027	8,38717555	19,8903505	132467,886	0,01863269	3844,48056	1,8688078
Glückwunsch	general	frequency	herzlich	345435	10,1966402	425,949991	213128158	0,25756486	2528377,33	25,758651
Haar	general	high	schulterlanges	1134978	10,8357097	50,9622946	4748675,83	0,83879474	37251,7625	83,925113
Haar	general	low	schmutzig	1134978	0,97831931	3,84599579	29,1576439	0,00044555	22,6697073	0,0904735
Idee	general	frequency	gute	3631331	3,32764958	423,812007	1838095,78	0,0125703	659778,482	1,3715340
Jahr	general	frequency	letztes	36081680	4,55019589	1260,08526	37878161,4	0,19113239	8340902,15	19,774985
Krankheit	general	high	unheilbare	1254709	4,6250431	38,8912976	37344,7342	0,00898625	7402,89669	0,9364011
Krankheit	general	low	schrekliche	1254709	10,0130491	83,3578509	7182810,04	0,39174299	86042,8714	39,211489
Lage	general	low	missliche	2578664	7,00599051	35,2063325	159451,453	0,10806406	9871,72543	10,891003
Lage	general	low	ausweglose	2578664	9,6288672	112,97653	10114198,7	0,67002744	157075,357	67,086085
Lärm	general	high	ohrenbetäubender	217206	12,7891746	58,1295538	23918933,9	0,27482863	54525,6718	27,486378
Lärm	general	low	unnötiger	217206	6,28093757	25,9863485	52525,0548	0,00298147	4681,18321	0,3019555
Leben	general	frequency	echtes	8671080	2,51912263	121,451429	84901,985	0,01109844	40196,5061	1,3420076
Linie	general	frequency	erste	1553632	5,16495057	770,594513	21697081,3	0,03973623	3781844,65	4,0179654
Mal	general	frequency	erstes	4145949	4,6162018	1002,30749	25136975,3	0,06813003	5466222,66	6,9816799
Meinung	general	frequency	eigene	2784440	3,50846801	234,163235	631298,284	0,00602833	207893,508	0,6536528
Netzwerk	general	frequency	sozial	1275074	7,19464428	488,481877	35077351,5	0,09086256	2074240,25	9,1222976
Papier	general	high	zusammengeknülltes	873123	9,46816169	10,801382	82649,8567	0,13413883	1318,53813	13,432835
Papier	general	low	zerknittertes	873123	7,97071281	12,5592541	39576,3097	0,04738682	1447,94996	4,7576301
Problem	general	frequency	großes	6869925	2,60088263	406,042692	1023723,9	0,01045337	481800,883	1,2253712
Quelle	general	frequency	andere	1571190	2,1414915	141,270114	89712,2374	0,00158727	49143,8152	0,2015457
Reihenfolge	general	high	alphabetische	266474	6,72867074	59,3681818	373958,776	0,0127808	26549,8826	1,2898546
Reihenfolge	general	low	beliebiger	266474	11,7709742	106,430797	39592813,1	0,42492634	169171,802	42,503562
Schule	general	high	weiterführende	4321027	-1,3451901	-9,74384246	37,4082778	-0,0005545	48,6964778	0,0359841
Schule	general	low	naturwissenschaftliche	4321027	8,10212235	229,42057	14479167,4	0,2503527	507755,17	25,12094
Seite	general	frequency	andere	8954459	3,6224444	738,676561	6865170,91	0,03652714	2212993,01	3,9038523
Tag	general	frequency	ganzer	14738939	3,44772705	534,425063	3154512,44	0,04545458	1064633,7	4,9661068
Türe	general	high	verschlossene	1593032	9,3615144	22,7469387	340478,495	0,22079776	5823,89463	22,113336
Türe	general	low	verriegelte	1593032	10,2911811	179,269746	40285647,5	0,4209216	414359,74	42,122274
Wandel	general	high	demographischer	483459	10,6746545	197,661706	63884462	0,43973156	527704,913	43,995816

Wandel	general	low	gesellschaftlicher	483459	5,86844182	118,101503	815904,067	0,01547154	90129,1569	1,5725236
Wasser	general	high	sprudelndes	4004467	6,6028804	35,9580192	125790,576	0,08508248	9586,48144	8,5965483
Wasser	general	low	klares	4004467	4,59529234	156,336164	592123,914	0,020522	119773,494	2,1378630
Woche	general	frequency	letzte	7093123	4,9370403	709,261995	15589912,6	0,05992627	2836044,22	6,1357627
Zeit	general	frequency	lange	16945136	4,92586476	902,39393	25032922,8	0,18147739	4558540,22	18,669798

Appendix 4.1. Stimuli sentences used in experiment IV

Sentence	Word1	Word2	Word3	Word4	Word5	Probe
Zusammenleben vor der Ehe ist in der **katholischen Kirche** eigentlich nicht erlaubt.	christlich	reformiert	orthodox	römisch	gotisch	heilig
Im vorliegenden Fall sind aber **mildernde Umstände** übermächtig.	widrig	äußere	ungeklärt	ungünstig	besondere	schwierig
Dieser Nachweis bescheinigt, wie und wo Strom aus **erneuerbaren Energien** produziert wurde.	regenerativ	elektrisch	kinetisch	erzeugt	sauber	positiv
Ich denke, dass die deutsche **gesetzliche Regelung**, die jetzt gefunden wurde, ein guter Kompromiss ist.	vertraglich	abweichend	geltend	getroffen	einheitlich	rechtlich
Wir sind sehr froh, dass die **freiwillige Feuerwehr** so schnell eingegriffen hat.	örtlich	eingesetzt	umgehend	zuständig	städtisch	eingesetzt
Die Pros und Kontras der **vorliegenden Verordnung** sind hinlänglich bekannt.	ärztlich	erlassen	geltend	ergangen	revidiert	geltend
Sie hatte vor längerer Zeit eine Art **epileptischen Anfall** erlitten und ist seitdem sehr vorsichtig.	hysterisch	akut	krampfartig	heftig	allergisch	akut
Ein Großteil der Kunstwerke ist in diversen **öffentlichen Einrichtungen** des Landes ausgestellt.	sozial	stationär	gemeinnützig	kulturell	wissenschaftlich	stationär

Ein Teil des Teams hat **langjährige Erfahrung** mit dieser Materie.	praktisch	positiv	schlecht	persönlich	wertvoll	bisherig
Eine unübertroffene Fülle an Beispielen für **staatliche Beihilfen** bietet das schöne Italien.	gewährt	unzulässig	unerlaubt	rechtswidrig	strafbar	einmalig
Diese Daten gehören nicht zu den **personenbezogenen Daten**, sondern sind anonymisiert.	gespeichert	technisch	sensible	erhoben	relevant	gesammelt
Im Brand- oder Gefahrenfall sind die Fluchtwege von **entscheidender Bedeutung** für Ihre Sicherheit.	besondere	zunehmend	zentral	groß	wachsend	enorm
Hier lesen Sie **wichtige Informationen** und aktuelle Neuigkeiten zu unseren Produkten.	ausführlich	detailliert	weiterführend	allgemein	zusätzlich	weiterführend
Die Zahl der jährlich von Jugendlichen **verübten Straftaten** wurde mit über 20 000 beziffert.	begangene	geschehene	unaufgeklärte	schändliche	gemachte	schändlich
Die im Lavendel enthaltenen **ätherischen Öle** wirken entspannungsfördernd auf verschiedene Bereiche des Körpers.	pflanzlich	heiß	duftend	pflegend	kostbar	duftend
Der Rat der **Europäischen Gemeinschaft** hat bestimmte Umsetzungsmaßnahmen getroffen.	international	religiös	häuslich	verschworen	dörflich	religiös
Zunächst wurden die Anwesenden über den **aktuellen Stand** des Förderantrags informiert.	derzeitig	jetzig	heutig	technisch	schwer	stabil
Es geht bei den **vorgesehenen Maßnahmen** vor allem um die Förderung der Waldwirtschaft.	geeignet	baulich	konkret	notwendig	erforderlich	präventiv
Gute Quellen für mehrfach **ungesättigte Fettsäuren** sind Lein- und Distelöl sowie Fisch.	vertraglich	lebenswichtig	wertvoll	ungesund	pflanzlich	gesättigt

Wir treten für die **soziale Sicherheit** unserer Mitarbeitenden ein und entlohnen sie leistungsbezogen.	innere	national	öffentlich	nötig	absolut	geprüft
Das Ziel ist es, eine **nachhaltige Entwicklung** zu erreichen.	wirtschaftlich	positiv	demographisch	aktuell	zukünftig	aktuell
Die Haftentlassung kommt daher im Lichte der **angeführten Rechtsprechung** nicht in Betracht.	höchstrichterlich	gefestigt	ergangen	ständig	bisherig	bisherig
Der Begriff **geldwerter Vorteil** ist auch als Sachleistungen oder Sachbezüge bekannt.	steuerlich	wesentlich	wirtschaftlich	erheblich	zahlreich	erheblich
Bis Ende 2002 soll die Regelung in **nationales Recht** umgesetzt werden.	geltend	bürgerlich	öffentlich	demokratisch	anwendbar	öffentlich
Das Gelände ist **öffentlicher Raum** und kann somit von jedem genutzt werden.	ländlich	deutschsprachig	eng	geschlossen	leer	breit
Ein vor dem Tod **ergangenes Urteil** kann nicht post mortem rechtskräftig werden.	angefochten	rechtskräftig	erstinstanzlich	angegriffen	vernichtend	verkündet
Ingesamt wachsen **im botanischen Garten** über 2000 in Australien heimische Pflanzen.	englisch	tropisch	heimisch	blühend	mediterran	traumhaft
Der BUND bezeichnet die **vorgeschlagenen Maßnahmen** als ungenügend.	geeignet	konkret	notwendig	erforderlich	geplant	wirksam
Es gibt ja kein **schlechtes Wetter**, sondern nur falsch angezogene Menschen.	schön	sonnig	herrlich	regnerisch	trocken	trocken
Sie befürchten, manche Straftäter würden nach **verbüßter Strafe** noch schwerere Verbrechen begehen.	gerecht	drakonisch	verhängt	drastisch	saftig	gerecht
Du kannst dich aus **dieser misslichen Lage** nur selbst befreien.	ruhig	zentral	geographisch	finanziell	schwierig	schwierig

Die Beklagte verteidigt das **angefochtene Urteil** und verweist auf ihr bisheriges Vorgehen.	rechtskräftig	erstinstanzlich	ergangen	vernichtend	angegriffen	vernichtend
Ich habe fast ein **schlechtes Gewissen**, dass es mir so gut geht.	ruhig	rein	gut	moralisch	sozial	böse
Einen Rechtsfehler dieser **tatrichterlichen Würdigung** zeigt die Revision nicht auf.	verständig	kritisch	rechtlich	gebührend	zutreffend	angemessen
Sie verfügt über ein **fotografisches Gedächtnis** und kann logisch denken.	kollektives	kulturelles	visuelles	prozedurales	implizites	gesellschaftliches
Die Würde des Menschen ist nach **geltender Rechtsprechung** die wichtigste Maxime des Grundgesetzes.	höchstrichterlich	bundesgerichtlich	gefestigt	ergangen	bisherig	vorherige
Die Hauptprüfung sollte in **regelmäßigen Abständen** wiederholt werden.	zeitlich	ausreichend	gewiss	gering	kurz	zeitlich
Die Beklagte verteidigt die **angefochtene Entscheidung** und wiederholt die dargelegte Rechtsauffassung.	endgültig	getroffen	richtig	gerichtlich	politisch	getroffen
Aber es hat keinen Sinn, an einer **unerwiderten Liebe** festzuhalten.	wahr	verboten	unerfüllt	heimlich	romantisch	wahr
Der Text enthält **rechtliche Hinweise**, die nicht verbindlich sind.	wertvoll	wichtig	nützlich	konkret	allgemein	allgemein
Er studiert an der **technischen Universität** über neue Materialien.	medizinisch	katholisch	führend	bayerisch	renommiert	ausländisch
Wann das Landgericht den Termin zur **mündlichen Verhandlung** ansetzt, ist bislang nicht bekannt.	schriftlich	ausdrücklich	anstehend	sachlich	schwierig	langwierig

Für Deutschland kann der **demografische Wandel** zu einem großen Problem werden.	stetig	tiefgreifend	technologisch	gesellschaftlich	grundlegend	schützend
Die seit 2002 **geltende Fassung** regelt den Schutz der Arbeitnehmerinnen vor Passivrauchen am Arbeitsplatz.	gekürzt	überarbeitet	gültig	aktualisiert	endgültig	revidiert
Wir sind gespannt darauf, wie die **junge Generation** den Faden aufnimmt!	nachfolgend	künftig	kommend	nachwachsend	alt	künftig
Es gibt keinen **stichhaltigen Beweis** dafür, dass er gelogen hat.	entlastend	eindeutig	derartig	klar	berechtigt	eindeutig
Die verantwortungsvolle **genetische Veränderung** von Pflanzen sei nicht neu.	tiefgreifend	hormonell	strukturell	positiv	klimatisch	hormonell
Dem Lehrer wird das **rechtliche Gehör** im Hinblick auf eine definitive Entlassung gewährt.	intakt	fein	absolut	nachlassend	geschult	intakt
Das Kind benötigt **soziale Kompetenzen**, um ein verantwortungsvolles Mitglied der Gesellschaft zu werden	fachlich	interkulturell	kommunikativ	sprachlich	methodisch	emotional
Die bereits im Vorfeld **festgelegten Bedingungen** verhindern langwierige Verhandlungen.	klimatisch	optimal	schwierig	erschwert	verändert	familiär
In unserer **globalisierten Welt** spielt Internationalität eine zunehmende Rolle.	westlich	virtuell	digital	fremd	modern	virtuell
Jetzt plötzlich werden von der Regierung **rechtliche Bedenken** vorgetragen.	moralisch	ethisch	erheblich	berechtigt	ernsthaft	begründet
Danke für die **moderne Technik** , die unser Leben vereinfacht und schöner macht.	innovativ	ausgereift	digital	eingesetzt	verwendet	digital

Im Augenblick sind im **vorliegenden Fall** die Voraussetzungen für eine Verurteilung nicht gegeben.	entscheidend	schwer	konkret	deutlich	vorgelegt	konkret
In diesem Eltern-Seminar erhalten **alleinerziehende Mütter** Unterstützung verschiedenster Art.	stillend	leiblich	erziehend	werdend	berufstätig	jung
Das Gericht verurteilte den Angeklagten wegen **gefährlicher Körperverletzung** zu drei Jahren Haft.	fahrlässig	strafbar	begangen	schwer	absichtlich	strafbar
Deutschland hat eine **lange Tradition** des gesetzlich beförderten sozialen Ausgleichs.	alt	religiös	kulturell	spirituell	lebendig	bewährt
Ich glaube, hier hat jemand das Konzept **einer juristischen Person** nicht verstanden.	betroffen	natürlich	beteiligt	versichert	verletzt	betroffen
Wir wissen, daß uns ständig **magnetische Felder** umgeben, aber wir können sie nicht wahrnehmen.	elektrisch	fruchtbar	abgeerntet	benachbart	angrenzend	abgeerntet
Der Rechtsstreit hat **grundsätzliche Bedeutung** und eine höchstrichterliche Entscheidung dient der Weiterentwicklung des Rechts.	besonder	entscheidend	zunehmend	zentral	herausragend	wachsend
Im Yoga gibt es viele **verschiedene Formen** und Traditionen.	unterschiedlich	schriftlich	digital	heutig	vielfältig	heutig
Der private Besitz von Gold wurde unter **drakonische Strafen** gestellt.	gerecht	verhängt	empfindlich	drastisch	ausgesprochen	drastisch
Diese Flecken lassen sich mit **lauwarmen Wasser** und einem Wartungsspray entfernen.	kaltes	frisch	heiß	kochend	saubere	klar
Der Staat legte den **rechtlichen Rahmen** und die innere Ausgestaltung der Anstalten fest.	zeitlich	geschützt	rechtlich	angemessen	passend	rechtlich

Nach Diskussion entschied sich eine **große Mehrheit** für diesen Antrag.	überwältigend	absolut	übergroß	deutlich	knapp	klar
Die Kommission teilt dem betreffenden Mitgliedstaat die Ergebnisse der Bewertung mit.	ersuchend	beigetreten	ausstellend	teilnehmend	überschuldet	angehörend
Mit ihren blonden, **schulterlangen Haaren** sah sie unheimlich hübsch aus.	grau	dunkel	lockig	glatt	gefärbt	glatt
Der Austritt erfolgt durch **schriftliche Erklärung** gegenüber dem Landesvorstand.	ausführlich	logisch	verständlich	schlüssig	allgemein	plausibel
Der Artikel mit Foto wurde in der **aktuellen Ausgabe** für August veröffentlicht.	gedruckt	heutig	vorliegend	erscheinend	monatlich	monatlich
Die Ermittlungen sind bei der **zuständigen Staatsanwaltschaft** in Berlin anhängig.	ermittelnd	verantwortlich	tätige	beteiligt	betreffend	betreffend
Feiner Sandstrand, **kristallklares Wasser** und eine schöne Dünenlandschaft begeistern jeden Besucher.	still	rein	tief	fließend	trüb	verschmutzt
Das Landgericht hat die **sofortige Beschwerde** der Klägerin zurückgewiesen.	körperlich	staatsrechtlich	gesundheitlich	erhoben	typisch	erhoben
Das ist der Beginn praktizierter **politischer Bildung** in Deutschland.	beruflich	kulturell	schulisch	ökonomisch	akademisch	frühkindlich
Erst Ende November soll die **endgültige Entscheidung** bekannt gegeben werden.	erstinstanzlich	definitiv	offenbar	schnell	falsch	richtig
Als Katzenhaarallergie wird fälschlicherweise eine **allergische Reaktion** auf die Haare einer Katze bezeichnet.	chemisch	positiv	körperlich	gespannt	entzündlich	entzündlich

Dies erfordert **gegenseitige Anerkennung** und Respekt der beiden Sozialpartner füreinander.	staatlich	steuerlich	gesellschaftlich	gebührend	verdient	mangelnd
Experten mit viel **praktischer Erfahrung** unterstützen die Werkstatt telefonisch bei technischen Problemen.	langjährig	positiv	schlecht	negativ	bisherig	langjährig
Das Motiv des bereits **vorbestraften Täters** bleibt unklar.	unbekannt	mutmaßlich	maskiert	jugendlich	potenziell	unbekannt
Eigentlich war ich immer der Meinung**, koffeinfreier Kaffee** schmeckt anders als normaler.	gemahlen	gebrüht	lecker	heiß	duftend	aromatisch
Der Schuldige soll vor Gericht gestellt werden und eine **gerechte Strafe** erhalten.	drakonisch	verhängt	empfindlich	drastisch	saftig	saftig
Eine direkte Antwort auf die **gestellte Frage** gab es allerdings nicht.	offen	folgend	häufig	wichtig	zentral	spannend
Eine gegen seine Entscheidung **gerichtete Beschwerde** war als unbegründet zurückzuweisen.	körperlich	staatsrechtlich	gesundheitlich	erhoben	typisch	chronisch
Neben der Verpflegung sind ausgewählte **alkoholfreie Getränke** beim Büfett inklusive.	kühl	erfrischend	zuckerhaltig	lecker	süß	kühl
Die Staatsanwaltschaft prüft weiter, ob eine strafrechtliche Verfolgung notwendig ist.	gerichtlich	konsequent	grausam	gnadenlos	brutal	rassistisch
Egal welche **sportliche Aktivität** ihr in eurer Freizeit ausübt, es schadet euch nicht.	körperlich	vielfältig	kulturell	gemeinsam	kriminell	vielfältig
Das Landesarbeitsgericht änderte die **erstinstanzliche Entscheidung** und wies die Klage ab.	endgültig	getroffen	richtig	politisch	strategisch	getroffen

Plötzlich standen sie vor zwei **verschlossenen Türen** und konnten nur noch umkehren.	geöffnet	breit	schmal	knarrend	zweiflügelig	verriegelt
Die Legitimation kann sich einzig aus der **allgemeinen Bestimmung** ergeben.	gesetzlich	geltend	einschlägig	rechtlich	vorstehend	gesetzlich
Das offene und **persönliche Gespräch** ist der kürzeste Weg, um Probleme zu lösen.	intensiv	klärend	anregend	konstruktiv	vertraulich	informell
Kommt es zu keiner Einigung, bleibt der **angefochtene Beschluss** unwirksam.	einstimmig	richterlich	angegriffen	getroffen	mehrheitlich	ablehnend
Die meisten Menschen bekommen jedes **gesprochene Wort** aktiv mit.	geschrieben	lobend	klar	richtig	deutlich	lobend
Diese Rechtsauffassung ist nach den **getroffenen Feststellungen** unzutreffend.	tatrichterlich	bindend	gerichtlich	gesondert	angegriffen	besondere
Unsere Reisen bieten **unvergessliche Erlebnisse**, neue Bekanntschaften und besondere Begegnungen.	traumhaft	einmalig	besondere	kulinarisch	beeindruckend	besondere
Wir empfehlen generell, sich bei einem **gerichtlichen Verfahren** anwaltlich unterstützen zu lassen.	vereinfacht	vorliegend	anhängig	laufend	aufwändig	vorliegend
Viele Kriegskinder haben ihre **traumatischen Erlebnisse** unbewusst an eigenen Kinder weitervererbt.	einmalig	besondere	kulinarisch	beeindruckend	positiv	negativ
Auf dem Gebiet des **materiellen Rechts** sind nur Einzelfragen staatlich geregelt.	geltend	bürgerlich	öffentlich	national	gleich	geltend
Diese Veröffentlichung ist lediglich eine **allgemeine Information** und dient nicht zur Klärung individueller Fragestellungen.	ausführlich	wichtig	detailliert	aktuell	relevant	nützlich

Sie bittet ihren Chef, Recherchen zu vier kürzlich **verübten Morden** anstellen zu dürfen.	vorbestraft	engagiert	mutig	aktiv	offensichtlich	passiv
Bis 2020 erhöht sich die Anzahl der **pflegebedürftigen Menschen** um voraussichtlich 37 Prozent.	jung	alt	behindert	normal	betroffen	behindert
Das Gericht verurteilte den Mann wegen **versuchten Mordes** zu einer Freiheitsstrafe von dreieinhalb Jahren.	brutal	kaltblütig	begangen	verübt	ungeklärt	verübt

Appendix 4.2. Stimuli used in experiment IV and the linguistic predictors

Collocation	**CollocType**	**TP_Forward**	**TP_Backward**	**FreqModifier**	**FreqNoun**	**FreqColloc**	**NCD**	**CDV**
aktuelle Ausgabe	general	0,01171327	0,05388393	5749460	1249816	67345	511	2
aktueller Stand	general	0,02300112	0,08327204	5749460	1588096	132244	916	19
alkoholfreies Getränk	general	0,29516787	0,0249766	53434	631471	15772	606	2
alleinerziehende Mutter	general	0,40683452	0,00924072	53434	2706	25013	7651	42
allergische Reaktion	general	0,33756876	0,04381223	2954460	895024	39213	1044	48
allgemeine Bestimmung	terminology	0,00255004	0,01074401	2954460	701228	7534	421	1
allgemeine Information	terminology	0,01808588	0,00963662	2954460	5544890	53434	3099	2
angefochtene Entscheidung	terminology	0,21276459	0,00471876	62078	2799043	13208	3682	1
angefochtene Urteil	terminology	0,29311511	0,01624366	62078	1120191	18196	876	1
angefochtener Beschluss	terminology	0,08780889	0,00803206	473699	678655	5451	178	1
angefü□hrte Rechtsprechung	terminology	0,15343575	0,05073848	473699	286804	14552	25	1
ätherische Öle	general	0,76716092	0,10213765	102674	735527	75125	458	2
betreffende Mitgliedstaat	terminology	0,01405745	0,03241367	9290	205438	6659	0	0
botanischer Garten	general	0,59415237	0,03762488	9290	1621374	61004	1353	17
demografischer Wandel	general	0,4059484	0,08091276	26594	483459	39118	110	0
drakonische Strafe	terminology	0,35349839	0,00837499	449541	392120	3284	1238	0
endg□ültige Entscheidung	terminology	0,04040699	0,01068365	449541	2799043	29904	3682	26

entscheidende Bedeutung	terminology	0,03142087	0,01969475	29174	2344077	46166	1284	15
epileptischer Anfall	general	0,77650919	0,14715977	29174	80416	11834	410	18
ergangene Urteil	terminology	0,15729112	0,00373418	29174	1120191	4183	876	0
erneuerbare Energie	general	0,83109883	0,16382806	2913581	2280519	373613	1423	15
erstinstanzliche Entscheidung	terminology	0,07043943	0,00073418	247508	2799043	2055	3682	0
europäische Gemeinschaft	terminology	0,01088386	0,03957247	92507	801340	31711	376	3
festgelegte Bedingung	terminology	0,00577759	0,00112961	1036807	1265923	1430	1133	1
fotografisches Gedächtnis	general	0,0141719	0,00555445	20943	236027	1201	896	31
freiwillige Feuerwehr	general	0,18309989	0,17672778	239384	978901	172999	316	5
gefährliche Körperverletzung	general	0,01242372	0,16548047	764040	77840	12881	257	2
gegenseitige Anerkennung	terminology	0,01159596	0,01512513	764040	577582	8736	385	0
geldwerter Vorteil	terminology	0,63629852	0,00544407	132462	2447801	13326	1226	0
geltende Fassung	terminology	0,02184228	0,04650194	215885	314589	14629	224	0
genetisch Veränderung	general	0,21278782	0,03553502	215885	1433459	50938	980	5
gerechte Strafe	terminology	0,01080179	0,02104713	1430370	392120	8253	1238	72
gerichtete Beschwerde	terminology	0,00942157	0,00196494	1430370	635135	1248	490	0
gerichtliches Verfahren	terminology	0,08543698	0,01170683	81017	1575320	18442	439	2
gesetzliche Regelung	terminology	0,0597307	0,08454756	330594	1010520	85437	48	1
gesprochenes Wort	general	0,24185048	0,00444539	330594	4407714	19594	7382	17
gestellte Frage	general	0,1866761	0,00664778	113245	9283395	61714	8201	7
getroffene Feststellung	terminology	0,06111528	0,02138897	1722325	323578	6921	57	0
globalisierte Welt	general	0,40807521	0,00271798	6197580	6994527	19011	8116	2
große Mehrheit	general	0,00276132	0,09454457	6197580	735304	69519	190	10
grundsätzliche Bedeutung	terminology	0,00662709	0,00486929	6197580	2344077	11414	1284	0
junge Generation	general	0,01246212	0,06407773	393367	1205333	77235	2	20
juristische Person	terminology	0,14471473	0,01154856	393367	4929274	56926	4388	4
katholische Kirche	general	0,25891689	0,06473756	393367	3044168	197072	1561	61
koffeinfreier Kaffee	general	0,31208687	0,00078282	2118	844385	661	3430	17
kristalklares Wasser	general	0,35457791	0,00337323	38096	4004467	13508	5241	2
lange Tradition	general	0,00770671	0,07309835	38096	1011555	73943	677	1

langjährige Erfahrung	general	0,21108932	0,03801894	9594627	4059424	154335	2126	3
lauwarmes Wasser	general	0,37659762	0,00459886	731136	4004467	18416	5241	8
m□ündliche Verhandlung	terminology	0,17748269	0,08425569	260262	548236	46192	690	1
magnetisches Feld	general	0,10792029	0,0069672	86916	4004467	18416	5241	8
materielles Recht	terminology	0,02045373	0,00201063	86916	3312396	6660	6767	0
mildernder Umstand	terminology	0,59738372	0,00292434	5504	1124356	3288	1779	65
missliche Lage	general	0,57837642	0,00496226	22124	2578664	12796	3116	49
moderne Technik	general	0,02490899	0,0405336	22124	1840769	74613	796	23
nachhaltig Entwicklung	general	0,07942727	0,02270116	22124	4451535	101055	575	4
natioles Recht	terminology	0,02075904	0,00756945	2995425	3312396	25073	6767	0
öffentliche Einrichtung	terminology	0,01492944	0,02874903	260262	1702214	48937	590	14
öffentliche Raum	terminology	0,03036561	0,02792853	1207811	3563918	99535	3208	5
personenbezogene Daten	general	0,76746555	0,02733196	1207811	3528909	96452	1085	0
persönliches Gespräch	general	0,01302213	0,02591269	3277886	2243187	58127	2310	26
pflegebed□rftiges Mensch	general	0,31871227	0,00124838	125676	15226172	19008	9061	1
politische Bildung	general	0,01819677	0,05052081	4463709	1277335	64532	262	0
praktische Erfahrung	general	0,0293097	0,01544628	2139326	4059424	62703	2126	13
rechtliche Rahme	terminology	0,01165562	0,00263439	2139326	285869	10598	435	1
rechtliches Bedenken	terminology	0,00655367	0,00148125	909261	4022948	5959	14	0
rechtliches Gehör	terminology	0,01563577	0,10630248	909261	133741	14217	205	0
rechtliches Hinweis	terminology	0,01736575	0,00814151	909261	1939443	15790	1417	0
regelmäßiger Abstand	general	0,04043743	0,1034769	909261	880496	91111	698	9
schlechtes Gewissen	general	0,02548544	0,29038981	2253135	337808	98096	1124	322
schlechtes Wetter	general	0,0257998	0,07008498	2253135	1416937	99306	1096	60
schriftliche Erklärung	terminology	0,01129774	0,01116111	2253135	868462	9693	1665	7
schulterlanges Haar	general	0,37431615	0,00229079	2253135	1134978	2600	3770	1
sofortige Beschwerde	terminology	0,03581806	0,0143623	3849100	635135	9122	490	0
soziale Kompetenz	general	0,02239445	0,08933313	857959	846987	75664	122	14
soziale Sicherheit	terminology	0,00774145	0,01156779	974195	2261106	26156	3738	5
sportliche Aktivität	general	0,0602169	0,04746205	1116820	1235998	58663	657	1

staatliche Beihilfe	terminology	0,01097402	0,17146535	23329	71478	12256	162	3
ständige Rechtsprechung	terminology	0,00844881	0,06089525	23329	286804	17465	25	0
stichhaltiges Beweis	terminology	0,08152943	0,00277495	4547	685418	1902	3289	19
strafrechtliche Verfolgung	terminology	0,07058287	0,0389579	4547	206736	8054	265	7
tatrichterliche Wü□rdigung	terminology	0,22366395	0,01302093	4547	78105	1017	12	0
technische Universität	general	0,03207522	0,04510183	2749849	1955619	88202	674	5
traumatisches Erlebnis	general	0,23254583	0,01227196	2749849	966512	11861	322	17
unerwiderte Liebe	general	0,52516411	0,00045733	2749849	3148687	1440	5619	7
ungesättigte Fettsäure	general	0,70666746	0,29783692	2749849	79285	55015	1	0
unvergessliches Erlebnis	general	0,18611362	0,05692118	295599	966512	23614	322	3
verübte Straftat	terminology	0,06206021	0,00150133	121461	200994	419	218	0
verübte Mord	terminology	0,27497936	0,00084923	291074	358348	538	2942	0
verbüßte Strafe	terminology	0,00898415	0,01694727	8669	392120	333	1238	0
verschiedene Form	general	0,26501511	0,02020612	291074	4293611	72765	1704	19
verschlossene Tür	general	0,01912916	0,01553797	12152	1593032	32189	6135	0
versuchte Mord	terminology	0,04833314	0,00208464	12152	358348	5568	2942	0
vorbestrafte Täter	terminology	0,01514154	0,00032748	129600	561861	184	723	0
vorgesehene Maßhme	terminology	0,01354022	0,00209536	129600	2072676	4346	411	0
vorgesehene Maßhme	terminology	0,05114198	0,0031978	662641	2072676	6628	411	0
vorliegende Verordnung	terminology	0,0247751	0,04826925	662641	340113	16417	62	0
vorliegender Fall	terminology	0,12832288	0,01000485	662641	8499082	85032	7324	7
wichtige Information	general	0,01351235	0,02509698	10298723	5544890	139160	3099	87
zuständige Staatsanwaltschaft	terminology	0,00302976	0,01326228	1274030	291051	3860	250	0